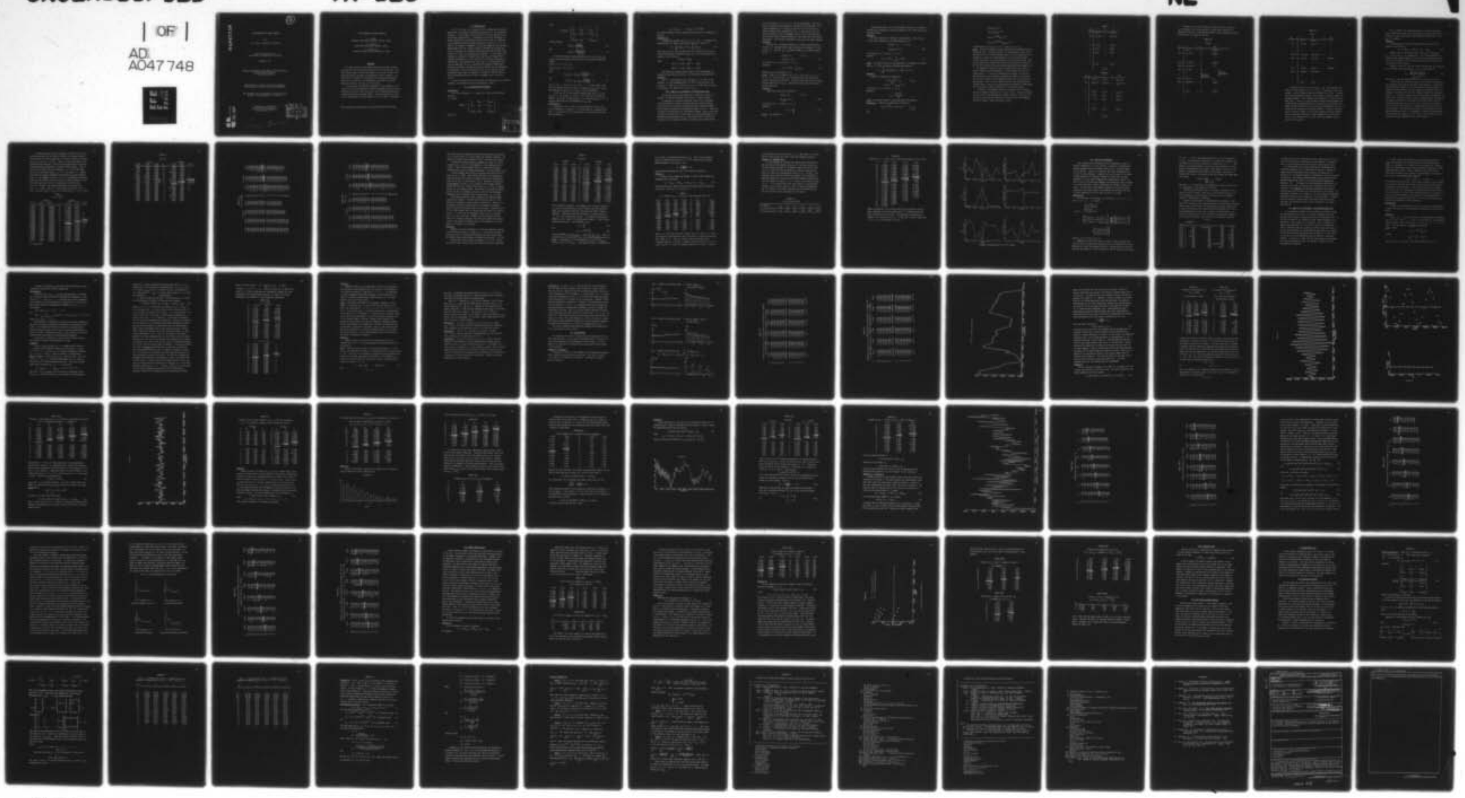
AD-A047 748 SOUTHERN METHODIST UNIV DALLAS TEX DEPT OF STATISTICS F/G 12/1
A NEW APPROACH TO ARMA MODELING.(U)
DEC 77 H L GRAY, G D KELLEY, D D MCINTIRE N00014-75-C-0439

UNCLASSIFIED TR-126 NL

| OF |
AD
A047 748



END
DATE
FILMED
1 -78
DDC

(12)

A NEW APPROACH TO ARMA MODELING

by

H. L. Gray, G. Kelley, D. McIntire

Technical Report No. 126
Department of Statistics ONR Contract

December, 1977

Research sponsored by the Office of Naval Research
Contract N00014-75-C-0439
Project NR 042-280

DEPARTMENT OF STATISTICS
Southern Methodist University
Dallas, Texas 75275

(See 1473)

A NEW APPROACH TO ARMA MODELING†

H. L. Gray
Southern Methodist University, Dallas, Texas

G. D. Kelley
Texas Tech University, Lubbock, Texas

D. D. McIntire
Anderson-Clayton Foods, Richardson, Texas

ABSTRACT

In recent years the Box-Jenkins method has become a popular technique for forecasting future behavior of a time series. Once the forecast model is known the method is very easy to employ and adequate computer packages are available for most purposes. Unfortunately the problem of determining the appropriate forecast model has, for models of any complexity, been one of the major stumbling blocks to the user of this method.

In this paper a satisfactory solution to that problem is obtained and it is demonstrated by numerous examples how this greatly enlarges the class of data sets which can be adequately modeled by autoregressive-moving average models. This new approach is sufficiently unequivocal that most users will find it easy to implement.

†This research was sponsored by ONR Contract #N00014-75-C-0439.

I. INTRODUCTION

In [7] H.L. Gray, A.G. Houston and F.W. Morgan have studied in some detail a new spectral estimator referred to as the G-Spectral Estimator. In that paper they pointed out that the problem of selecting the appropriate G_n - transform for spectral estimation is almost equivalent to the problem of selecting p and q in modeling an autoregressive-moving average process of order (p, q), i.e. an ARMA (p, q) process. Moreover, it was also suggested and implicitly demonstrated that the "R and S arrays" (defined later) which are used to calculate the G-Spectral estimator could be utilized to determine p, d and q in an ARIMA (p,d,q) process. In this paper that suggestion is explored in depth and a new set of criteria which uniquely determines p, d and q and which is readily observable is established. Many examples are given and it is demonstrated that this new approach not only simplifies the problem of ARMA model fitting, but greatly extends the range of models that can reasonably be identified. This is particularly true in the case of the nonstationary model where the method is utilized to identify and remove nonstationarities due to real or complex roots on the unit circle.

In the following section we set down the definitions and theorems necessary to establish the results referred to above.

II. DEFINITIONS AND THEOREMS

Definition 1.

Let m be an integer, $h > 0$, and f be a real valued function. Also let

$$f_m = f(mh),$$

$$H_n[f_m] = \begin{vmatrix} f_m & f_{m+1} & \cdots & f_{m+n-1} \\ f_{m+1} & f_{m+2} & \cdots & f_{m+n} \\ \vdots & \vdots & & \vdots \\ f_{m+n-1} & f_{m+n} & & f_{m+2n-2} \end{vmatrix},$$

$$H_0[f_m] \equiv 1, \qquad (1)$$

and

$$H_n[1;f_m] = \begin{vmatrix} 1 & 1 & \cdots & 1 \\ f_m & f_{m+1} & \cdots & f_{m+n} \\ f_{m+1} & f_{m+2} & \cdots & f_{m+n+1} \\ \vdots & & & \\ f_{m+n-1} & f_{m+n} & \cdots & f_{m+2n-1} \end{vmatrix} .$$

Then we define

$$R_n(f_m) = \frac{H_n[f_m]}{H_n[1;f_m]} \tag{2}$$

and

$$S_n(f_m) = \frac{H_{n+1}[1;f_m]}{H_n[f_m]} . \tag{3}$$

It has been shown by Pye and Atchison [10] that $R_n(f_m)$ and $S_n(f_m)$ can be calculated recursively by the following relations.

Define

$S_0(f_m) = 1$, $m = 0, \mp 1, \mp 2, \ldots$

$R_1(f_m) = f_m$, $m = 0, \mp 1, \mp 2, \ldots$.

Then

$$R_{n+1}(f_m) = R_n(f_{m+1})\left[\frac{S_n(f_{m+1})}{S_n(f_m)} - 1\right] \tag{4}$$

and

$$S_n(f_m) = S_{n-1}(f_{m+1})\left[\frac{R_n(f_{m+1})}{R_n(f_m)} - 1\right] \tag{5}$$

for $n = 1,2, \ldots$ and $m = 0, \mp 1, \mp 2, \ldots$. In the future, when they are defined we use (4) and (5) to define $R_{n+1}(f_m)$ and $S_n(f_m)$ even though the ratio's in (2) and (3) may be undefined.

In (4) and (5) we have tacitly assumed that $S_n(f_m) \neq 0$ and $R_n(f_m) \neq 0$. If $R_n(f_{m+1}) = R_n(f_m) = 0$ we leave $S_n(f_m)$ undefined. However, if $R_n(f_m) = 0$ and $R_n(f_{m+1}) \neq 0$ we define $S_n(f_m) = \mp \infty$. A similar definition will be employed for $R_{n+1}(f_m)$.

<u>Definition 2.</u>

A function f will be said to be an element of $L(n,\Delta)$ over a set of integers $I = \{m_0, m_0+1, \ldots\}$ if there exists a smallest integer $n > 0$ and a set of ϕ_i's such that f is a solution of the difference equation

$$y_m - \phi_1 y_{m-1} - \ldots - \phi_n y_{m-n} = 0 \text{ for } m \in I.$$

For future reference we now include the definition of an ARMA (p, q) process.

Definition 3.

A stochastic process $\{X_t\}$, $t = 0, \mp 1, \mp 2, \ldots$ is said to be autoregressive of order p and moving average of order q if

$$X_t = \sum_{k=1}^{p} \phi_k X_{t-k} + Z_t - \sum_{k=1}^{q} \theta_k Z_{t-k}, \tag{6}$$

where the ϕ_k and θ_k are constants and Z_t is white noise. By utilizing the operator B defined by $BX_t = X_{t-1}$ we will usually write equation (6) in the form

$$\phi(B)X_t = \theta(B)Z_t \tag{7}$$

where

$$\phi(B) = 1 - \phi_1 B - \phi_2 B^2 \ldots - \phi_p B^p$$

$$\theta(B) = 1 - \theta_1 B - \theta_2 B^2 \ldots - \theta_q B^q.$$

The following theorem forms a basis for the development of several of the results to follow. Its proof can be found in [7].

Theorem 1.

Let $n > 0$ and suppose $S_n(f_m)$ and $R_n(f_m)$ in (2) and (3) are defined, i.e. the denominators in (2) and (3) are not zero, and $S_n(f_m) \neq 0$. Then $f_m \in L(n,\Delta)$ for $m \geq m_0 + n$ if and only if $S_n(f_m)$ is constant, as a function of m, for $m \geq m_0$.

III. ARMA (p, q) Processes - The Stationary Case

The current most popular approach to modeling an ARMA (p, q) process is a method referred to as the Box - Jenkins method.

It consists of initially examining graphs of the estimated autocorrelation and partial autocorrelation to obtain possible values of p and q and then investigating these contending models more closely to determine an appropriate model, see [5]. There are a number of difficulties with the method, however most of them stem from the fact that even if the graphs of the autocorrelation, $\rho(\tau)$, and partial autocorrelation, ϕ_{kk}, were error free the simple inspection of their graphs would not in general yield unique values

of p and q when $p > 0$, $q > 0$, i.e. in the mixed model. When this fact is coupled with the ambiquity induced by the necessity of utilizing estimators $\hat{\rho}(\tau)$ and $\hat{\phi}_{kk}$ in place of the true values the method becomes richly infused with art. In this section we will show how Theorem 1 can be expanded in such a way as to establish new criteria which uniquely determine p and q and which are readily observable. The following theorems are the basis for these remarks.

Theorem 2.

Let $\{X_t\}$ be a stationary ARMA (p, q) process with autocorrelation $\rho(\tau)$. Suppose that $S_n(f_m)$ and $R_n(f_m)$ are defined $p > 0$ and $S_n(f_m) \neq 0$. Then for $f_m = \rho_m$, some integer m_0, and some constant $C_1 \neq 0$

$$\begin{aligned} S_n(f_m) &= C_1 \quad , m \geq m_0 \\ S_n(f_{m_0-1}) &\neq C_1 \end{aligned} \tag{8}$$

if and only if $n = p$ and $m_0 = q - p + 1$.

Moreover

$$C_1 = (-1)^p [1 - \sum_{k=1}^{p} \phi_k], \tag{9}$$

where the ϕ_k's are defined in (7).

Proof. Since $\{X_t\}$ is ARMA (p, q), $\rho(\tau) \in L(p,\Delta)$ for $\tau \geq q + 1 = q - p + 1 + p$ but not for $\tau \geq q$ and hence the first part of the theorem follows from Theorem 1. From the first part of this theorem and the definition of $S_n(\rho_m)$, equation (9) then follows by simple column additions.

Theorem 3.

Under the conditions of Theorem 2

$$\begin{aligned} S_n(\rho_m) &= C_2 \quad , m \leq m_1 \\ S_n(\rho_{m_1+1}) &\neq C_2 \end{aligned} \tag{10}$$

if and only if $n = p$ and $m_1 = -q - p$.

Moreover

$$C_2 = -\frac{C_1}{\phi_p} . \tag{11}$$

Proof. See Appendix 1.

Although Theorems 2 and 3 are extremely useful, the following theorems will probably be even more helpful in the data that will be considered later.

Theorem 4.

Let the conditions of Theorem 2 be satisfied. Then for $f_m = (-1)^m \rho_m$, some integer m_0, and some constant $C_3 \neq 0$

$$S_n(f_m) = C_3 \quad , \quad m \geq m_0$$

$$S_n(f_{m_0-1}) \neq C_3 \tag{12}$$

if and only if $n = p$ and $m_0 = q - p + 1$.

Moreover

$$C_3 = (-1)^p (1 - \sum_{k=1}^{p} (-1)^k \phi_k). \tag{13}$$

Proof. The proof follows in the same manner as Theorem 2 by noting that $f_m = (-1)^m \rho_m$ satisfies the difference equation

$$(1 - \sum_{k=1}^{p} (-1)^k \phi_k B^k) f_m = 0, \text{ when } m \geq q + 1.$$

Theorem 5.

Under the conditions of Theorem 4

$$S_n((-1)^m \rho_m) = C_4 \qquad m \leq m_1$$

$$S_n((-1)^{m_1+1} \rho_{m_1+1}) \neq C_4$$

if and only if $n = p$ and $m_1 = -q - p$.

Moreover

$$C_4 = \frac{(-1)^{p+1} C_3}{\phi_p} .$$

Proof. The proof follows in the same manner as Theorem 3.

Corollary 1. Under the assumptions of Theorem 2

$$R_{n+1}((-1)^m \rho_m) = R_{n+1}(\rho_m) = 0 \qquad \begin{array}{l} m \geq m_0 \\ m \leq m_1, \end{array}$$

and

$$R_{n+1}(\rho_{q-p}) \neq 0$$

$$R_{n+1}(\rho_{-q-p+1}) \neq 0$$

$$R_{n+1}((-1)^{q-p}\rho_{q-p}) \neq 0$$

$$R_{n+1}((-1)^{-q-p+1}\rho_{-q-p+1}) \neq 0$$

if and only if $n = p$, $m_0 = q - p + 1$ and $m_1 = - q - p$.

Proof. The result follows at once from the above theorems and the recursion rule. Although Corollary 1 characterizes p and q theoretically as well as our previous results, we will seldom use it to determine p since its column behavior is not as distinct when errors are present as that described by our previous results. Thus, p will normally be identified when errors are present by taking p as the first column having the behavior described by Theorems 2 and 3. In contrast we shall see that the behavior of the R array is usually adequate for determining q. These remarks will all be clear shortly. In order to demonstrate the manner in which the above results can be utilized to identify an ARMA (p, q) process we will now consider several examples making use of the true autocorrelation. This is not to imply that the true autocorrelation is ever known or used in an actual problem but simply to demonstrate the true state of nature as pertains to the above results.

In each of the examples which are to follow we will display arrays of numbers referred to as R and S arrays. They are computed via (4) and (5) and arranged as in Tables I and II. Since f_m will be given in each case we shorten the notation within the tabular values to $R_n(f_m) = R_n(m)$ and $S_n(f_m) = S_n(m)$.

TABLE I

$m \backslash n$	1	...	k
$-\ell$	$S_1(-\ell)$	...	$S_k(-\ell)$
$-\ell+1$	$S_1(-\ell+1)$	...	$S_k(-\ell+1)$
.	.		
.	.		
.	.		
-1	$S_1(-1)$	...	$S_k(-1)$
0	$S_1(0)$	...	$S_k(0)$
1	$S_1(1)$	...	$S_k(1)$
2			
.	.		
.	.		
.	.		
j	$S_1(j)$	...	$S_k(j)$

$S_n(f_m)$

TABLE II

$m \backslash n$	1	2		k+1
$-\ell$	$R_1(-\ell)$	$R_2(-\ell)$	...	$R_{k+1}(-\ell)$
$-\ell+1$	$R_1(-\ell+1)$	$R_2(-\ell+1)$	...	$R_{k+1}(-\ell+1)$
.	.	.		
.	.	.		
.	.	.		
-1	$R_1(-1)$	$R_1(-1)$	...	$R_{k+1}(-1)$
0	$R_1(0)$	$R_1(0)$	...	$R_{k+1}(0)$
1	(1)	$R_1(1)$	...	$R_{k+1}(1)$
2	.	.		.
.	.	.		.
.	.	.		.
j	$R_1(j)$	$R_1(j)$		$R_{k+1}(j)$

$R_n(f_m)$

In terms of the R and S arrays of Tables I and II (with $f_m = \rho_m$), Theorems 2 and 3 and Corollary 1 state that a process is a stationary ARMA (p, q) process if and only if the associated R and S arrays are as in Tables III and IV.

TABLE III

$(S(\rho_m))$

$m \backslash n$	1	...	p	p+1
$-\ell$	$S_1(-\ell)$	...	C_2	$C_2[\frac{0}{0}-1]$
$-\ell+1$	$S_1(-\ell+1)$		C_2	$C_2[\frac{0}{0}-1]$
.	.		.	.
.	.		.	.
.	.		.	.
$-q-p-2$	$S_1(-q-p-2)$		C_2	$C_2[\frac{0}{0}-1]$
$-q-p-1$	$S_1(-q-p-1)$	...	C_2	$\mp\infty$
$-q-p$	$S_1(-q-p)$	...	C_2	
.	.			2q non-constants
.	.		2q non-constants	
.	.			
$+q-p$	$S_1(q-p)$	...		$-C_1$
$q-p+1$	$S_1(q-p+1)$	...	C_1	$C_1[\frac{0}{0}-1]$
.	.		.	.
.	.		.	.
.	.		.	.
j	$S_1(j)$	...	C_1	$C_1[\frac{0}{0}-1]$

TABLE IV

$R(\rho_m)$

$m \backslash n$	1	2	...	p+1
$-\ell$	$\rho(-\ell)$	$R_2(-\ell)$	...	0
$-\ell+1$	$\rho(-\ell+1)$	$R_2(-\ell+1)$	...	0
.	.	.		.
.	.	.		.
.	.	.		.
$-q-p-1$	$\rho(-q-p-1)$	$R_2(-q-p-1)$	...	0
$-q-p$	$\rho(-q-p)$	$R_2(-q-p)$	...	Nonzero
.	.	.		.
.	.	.		.
.	.	.		.
$q-p$	$\rho(q-p)$	$R_2(q-p)$	...	Nonzero
$q-p+1$	$\rho(q-p+1)$	$R_2(-q-p+1)$	...	0
.	.	.		.
.	.	.		.
.	.	.		.
j	$\rho(j)$	$R_2(j)$	...	0

In Tables III and IV we have let $f_m = \rho_m$. The arrays do not differ in their pattern if $f_m = (-1)^m \rho_m$. Of course, in that case Column 1 in the R-array is made up of values of $(-1)^m \rho_m$. In Table III we have included Column p+1 even though it contains several undefined elements. Such elements are indicated by a factor of the form $[\frac{0}{0}-1]$. That is, in the presence of noise we would expect the column having the characteristics of Column p to be followed by a highly variable column since in practice the $\frac{0}{0}$ quantity will be replaced by the quotient of two small numbers. A similar consideration of the p+2 column in the R-array suggests the zero behavior should essentially continue. Later examples will clarify this. Also in later examples we will star the array elements $S_m(-m)$, $S_m(-m+1)$ and $R_{m+1}(-m)$.

This is because the constancy behavior in the S-array must occur symetrically about one of these pairs of points and the zero behavior in the R-array must occur symetrically about one of the $R_{m+1}(-m)$ elements.

Example 1.

To demonstrate the previous theorems consider the R and S arrays for the process

$$X_t - 1.32X_{t-1} + .68X_{t-2} = Z_t - .8Z_{t-1}. \tag{14}$$

Tables V and VI show these arrays when ρ_m is known using $f_m = (-1)^m \rho_m$ and $f_m = \rho_m$ respectively. It is well worthwhile to inspect those tables closely now since all of the previous results are well illustrated there and the remaining part of the paper is predicated on those results.

Tables VII and VIII show the R and S arrays when ρ_m is replaced by the sample autocorrelation $\hat{\rho}_m$, i.e. ρ_m is replaced by

$$\hat{\rho}_m = \frac{\sum_{i=1}^{T-m} (X_i - \bar{X})(X_{i+m} - \bar{X})}{\sum_{i=1}^{T} (X_i - \bar{X})^2}, \tag{15}$$

where the X_i are data simulated from (14) with the $Z_t \sim N(0,1)$.

In each case the arrays are shown to at least one column more than necessary and several rows more than necessary. In practice some preassigned number of rows and columns must be selected. Selecting this much larger than one expects is necessary induces no problem. For Tables VII and VIII a sample of size 200 was used to obtain $\hat{\rho}_m$. One should realize that in calculating the elements of the R and S arrays the number of lags utilized in these calculations increases as you move to the right of the table and as you move vertically away from the $S_i(-i+1)$, $S_i(-i)$ and $R_{i+1}(-i)$ elements. For this reason the values more removed from these points will be more variable. Fortunately the values in the proximity of these points determine p and q. In Appendix 2 a table is given showing the smallest and largest lag (i.e. value of m) utilized in calculating individual elements of the R and S arrays.

In observing arrays such as presented in Tables VII and VIII several approaches can be taken to identify the pattern present. To these authors the best approach is to first inspect the S-array with $f_m = \hat{\rho}_m$ and $f_m = (-1)^m \hat{\rho}_m$. Choose the one which appears most distinctive and identify p as the first column having the correct constant behavior followed by a highly variable column. This is very clear in Table VII where obviously p = 2. From that same table it is also clear that q - p + 1 = 0 so that q = 1. Notice that there are several different places where one could identify p and q. For example the fact p = 2, q = 1 is obvious at a glance since the two starred values in $S_2(m)$ are clearly distinct. Also note this choice of p and q is confirmed by the fact that $|S_3(-4)| = 44.4542$, the largest magnitude in the table, and $S_3(-1) = -3.5896$. This is of course significant since if p = 2, q = 1, the true value of $|S_3(-4)| = \infty$ while $S_3(-1) = -C_1 = -S_2(0)$.

TABLE V

$$f_m = (-1)^m \rho_m$$

R-Array				S-Array		
$R_1(m)$	$R_2(m)$	$R_3(m)$	m	$S_1(m)$	$S_2(m)$	$S_3(m)$
.13829	.0378	.000000	-10	-2.2806	4.4117	
-.17710	-.0553	.000000	-9	-1.7928	4.4117	.
.14042	.1693	.000000	-8	-1.0863	4.4117	.
-.01212	.2224	.000000	-7	14.0844	4.4117	.
-.18295	-.1005	.000000	-6	-3.0386	4.4117	u†
.37298	.0992	.000000	-5	-2.2198	4.4117	u
-.45498	-.1530	.000000	-4	-1.7356	4.4117	∞
.33468	.5636	.143284	-3	-.9420	4.4117	6.2513*
.01940	.4758	.299323*	-2	-28.3076	5.7403*	-1.5639*
-.52985	-.4701*	.074883	-1	-2.8873*	2.0857*	-3.0000
1.00000*	.1708	.000000	0	-1.5298*	3.0000	u
-.52985	-.3235	.000000	1	-1.0366	3.0000	u
.01940	-.3832	.000000	2	16.2492	3.0000	u
.33468	.1040	.000000	3	-2.3594	3.0000	.
-.45498	-.0674	.000000	4	-1.8197	3.0000	.
.37298	.0683	.000000	5	-1.4905	3.0000	.
-.18295	-.1512	.000000	6	-.9337	3.0000	
-.01212	-.1151	.000000	7	-12.5774	3.0000	
.14042	.0376	.000000	8	-2.2612	3.0000	
-.17710	-.0257	.000000	9	-1.7808	3.0000	

†u = underfined

TABLE VI

$f_m = \rho_m$

R-Array				S-Array		
$R_1(m)$	$R_2(m)$	$R_3(m)$	m	$S_1(m)$	$S_2(m)$	$S_3(m)$
.1382	-.3078	.0000	-10	.2806	.5294	
.1771	.4789	.0000	-9	-.2071	.5294	.
.1404	.2014	.0000	-8	-.9136	.5294	.
.0121	.1947	.0000	-7	-16.0844	.5294	.
-.1829	.2940	.0000	-6	1.0386	.5294	u
-.3729	1.0021	.0000	-5	.2198	.5294	u
-.4549	-1.0045	.0000	-4	-.2643	.5294	∞
-.3346	-.5018	-1.194030	-3	-1.0579	.5294	-3.203096*
.0194	-.5119	.973987*	-2	26.3076	1.7640*	-.801331*
.5298	-1.5298	-.243666	-1	.8873*	.6409*	-.360000
1.0000	.5558	.0000	0	-.4701*	.3600	u
.5298	.3481	.0000	1	-.9633	.3600	u
.0194	.3412	.0000	2	-18.2492	.3600	u
-.3346	.6831	.0000	3	.3594	.3600	u
-.4549	-.6814	.0000	4	-.1802	.3600	.
-.3729	-.1999	.0000	5	-.5094	.3600	.
-.1829	-.1324	.0000	6	-1.0662	.3600	.
.0121	-.1369	.0000	7	10.5774	.3600	
.1404	-.3256	.0000	8	.2612	.3600	
.1771	.2093	.0000	9	-.2191	.3600	

TABLE VII

$(f_m = (-1)^m \hat{\rho}_m)$

R-Array					S-Array		
$R_1(m)$	$R_2(m)$	$R_3(m)$	$R_4(m)$	m	$S_1(m)$	$S_2(m)$	$S_3(m)$
.2404	.0679	-.0138	.0229	-10	-2.2615	3.9279	4.4171
-.3032	-.0841	-.0271	.1038	-9	-1.7551	4.5711	.6872
.2290	.2619	-.0316	.0156	-8	-1.1107	4.0987	-1.5689
-.0253	.3654	-.0184	.1423	-7	10.3686	3.7441	-.2366
-.2881	-.1261	-.0174	.0569	-6	-2.7838	4.2895	1.7045
.5139	.1313	-.0253	-.2514	-5	-2.1009	3.7228	-2.1323
-.5657	-.1718	-.0127	-.1365	-4	-1.6131	4.2711	-44.4542
.3469	.7293	.1215	-.3159*	-3	-.8142	4.1969	5.4827*
.0644	.4241	.2449*	.0894	-2	-10.0291	5.3992*	-1.5889*
-.5818	-.4182*	.0710	.0190	-1	-2.7187*	2.2373*	-3.5896
1.0000*	.1733	-.0098	.0257	0	-1.5818*	3.1537	3.3574
-.5818	-.3187	-.0199	.0173	1	-1.1108	3.2507	-.9606
.0644	-.5551	-.0143	-.1352	2	4.3825	3.3675	.2076
.3469	.1554	-.0152	.0292	3	-2.6310	3.0586	2.0505
-.5657	-.0937	-.0240	-.0530	4	-1.9084	3.5558	-.4409
.5139	.1197	-.0203	-.0222	5	-1.5606	2.8430	-1.5926
-.2881	-.2535	-.0098	-.0123	6	-.9120	3.0704	-5.2171
-.0253	-.1759	.0060	-.0253	7	-10.0349	3.2407	5.5352
.2290	.0694	.0153	.0063	8	-2.3244	3.5189	-3.6087
-.3032	-.0668	-.0051		9	-1.7927	2.7119	.8652
.2404	.0731	-.0068		10	-1.2946	2.5241	
-.0708	4.4244			11	.0424	2.5202	
-.0738	-.0352			12	-2.5003		

TABLE VIII

$(f_m = \hat{\rho}_m)$

R-Array					S-Array		
$R_1(m)$	$R_2(m)$	$R_3(m)$	$R_4(m)$	m	$S_1(m)$	$S_2(m)$	$S_3(m)$
.2404	-.5872	-.1090	-.0655	-10	.2615	.4962	-1.5419
.3032	.6024	.3045	.0589	-9	-.2449	.4064	-1.2102
.2290	.3271	-.1651	.0314	-8	-.8893	.7847	-.7783
.0253	.3064	.1900	-.0371	-7	-12.3686	.3618	-.9070
-.2881	.4478	-.1444	-.0851	-6	.7838	.5154	-1.1402
-.5139	2.7360	.1929	.8414	-5	.1009	.4882	-.6371
-.5657	-.7162	-.1517	2.0942	-4	-.3869	.3567	2.8973
-.3469	-.5008	-1.0973	-.6579*	-3	-1.1858	.4647	-2.6322*
.0644	-.5298	.9264*	.1861	-2	8.0291	1.4273*	-.7628*
.5818	-1.5818*	-.2685	-.2913	-1	.7187*	.5914*	-.2340
1.0000*	.6555	-.0886	-.0858	0	-.4182*	.3492	-1.0032
.5818	.3981	.2388	-.0259	1	-.8892	.2715	-.6426
.0644	.3811	-.1087	.0353	2	-6.3825	.4416	-.7958
-.3469	.6479	.1267	.0588	3	.6310	.3675	-1.0172
-.5657	-1.9509	-.2484	-.0301	4	-.0916	.3436	-.7764
-.5139	-.4252	.1059	.0636	5	-.4394	.5443	-.5559
-.2881	-.2125	-.1098	-.2739	6	-1.0880	.2730	-.2337
.0253	-.2197	-.0471	-.0880	7	8.0349	.4094	-1.5920
.2290	-.4970	.1202	.0530	8	.3244	.4482	-.4273
.3032	.5775	.0254		9	-.2073	.5415	-1.3207
.2404	.1342	-.0267		10	-.7054	.6439	
.0708	.0919			11	-2.0424	.4571	
-.0738	.1758			12	.5003		
-.1107				13	-.2941		
-.0782				14			

Thus when one puts together the total picture implied by Theorems 2-5, no other choice of p and q seems even remotely possible. The validity of this latter statement will be even more clear later.

It should be recalled that $\hat{\rho}_m$ appears in Column 1 of the R-array when $f_m = \hat{\rho}_m$ and $(-1)^m\hat{\rho}_m$ appears in that column when $f_m = (-1)^m\hat{\rho}_m$. Thus in this example errors in the autocorrelation estimates can be seen by comparing Tables 5 or 6 with Tables 7 or 8.

In this example both p and q are easily seen from the R or S array at $f_m = (-1)^m\hat{\rho}_m$. As we shall see this is not always the case and in fact the S-array is usually better for determining p while the R-array is often better for q. Moreover, although <u>it is not always</u> the case, $f_m = \hat{\rho}_m$ usually yields better identification for very high frequency data while $f_m = (-1)^m\hat{\rho}_m$ is usually better for low frequency data. Since most of the data we examine will be of the low frequency nature, in most of our remaining examples we will consider R and S arrays only at $f_m = (-1)^m\hat{\rho}_m$.

Some remarks should be made concerning the R-array before leaving this example. Knowing from the S-array that p = 2 we direct our attention to $R_3(m)$ column of Table VII. Since p = 2, the number of moving average terms is the number of nonzeros above or below the starred value in that column. Since $R_3(-3)$ is 5 to 10 times larger than the values near and above it we have at once q = 1. Note that the assumption that $R_3(-3) \neq 0$ implies $R_3(-1) \neq 0$ and vice versa. Finally we should stress that the "zero" and "constant behavior" must take place about the starred values. Such behavior anywhere else in the arrays is not indicative of an ARMA (p, q) process.

<u>Example 2</u>.

Consider the series referred to in Box and Jenkins as Series D and identified there as an ARMA (1, 0). The series consists of 310 points and its estimated autocorrelations are obvious from the first column of the R-array in Table IX since $R_1(m) = (-1)^m\hat{\rho}_m$.

Clearly from either array the process is a (1, 0). Notice how closely these arrays follow the theoretical pattern described

TABLE IX

Series D

$R_1(m)$	$R_2(m)$	$R_3(m)$	m	$S_1(m)$	$S_2(m)$	$S_3(m)$
	R-Array				S-Array	
-.2703	-.0022	.0040	-9	-2.1470	-1.7608	-.0452
.3101	-.0041	.0102	-8	-2.1315	-.0295	-1.8241
-.3509	-.0041	-.4220	-7	-2.1563	.0429	-4.5033
.4057	-.0041	-.0004	-6	-2.1343	4.5076	-50.2349
-.4602	.0044	.0046	-5	-2.1532	4.0960	-4.1915
.5307	-.0039	.0337	-4	-2.1712	-.6560	5.0495
-.6215	-.0051	.0801	-3	-2.1849	3.6692	60.0155*
.7365	.0035	.1355*	-2	-2.1697	86.8522*	-1.8438*
-.8615	-.1385*	.0042	-1	-2.1608*	1.9022*	-.2479
1.0000*	.0031	.0039	0	-1.8615*	4.5121	-3.4356
-.8615	-.0043	-.0276	1	-1.8549	.4281	-3.9221
.7365	-.0033	-.0004	2	-1.8440	3.9733	-39.8316
-.6215	.0038	.0036	3	-1.8539	3.6023	-.0260
.5307	-.0035	.0060	4	-1.8672	-.0383	-.5124
-.4602	-.0036	-.1154	5	-1.8816	.0254	-.8472
.4057	-.0036	.0001	6	-1.8648	.8468	-174.3048
-.3509	-.0020	-.0125	7	-1.8838	.8213	-7.2421
.3101	-.0011	-.0035	8	-1.8718	10.1191	-6.7227
-.2703	.0048	.0053	9	-1.8795	2.6900	-5.1405

by our previous theorems. For example, the "$-C_1$ and ∞ behavior" in $S_2(m)$ are as distinctive as the constant and zero behavior. Again the increased variability following the stable column is apparent. Thus again when one considers the entire picture, a (1, 0) is the only possible choice. Of course this particular example is also clear by the Box-Jenkins method. A final remark should be made here. From Theorems 3 and 5

$$\phi_p = -\frac{C_1}{C_2} \tag{16}$$

and

$$\phi_p = (-1)^{p+1}\frac{C_3}{C_4} \tag{17}$$

If we estimated C_1 by $S_p(\rho_{q-p+1})$ and C_2 by $S_p(\rho_{-q-p})$ then the resulting $\hat{\phi}_p$ is easily shown to be the Yule-Walker estimate of ϕ_p given q. Similarly if C_3 is estimated by $S_p((-1)^{q-p+1}\rho_{q-p+1})$ and C_4 by $S_p((-1)^{-q-p}\rho_{-q-p})$ the result is again the conditional

(i.e. given q) Yule-Walker estimate of ϕ_p. Thus in this example, since we simultaneously know p = 1, q = 0, we also have the Yule-Walker estimate of ϕ_1 given q, i.e.

$$\hat{\phi}_1 = \frac{1.8615}{2.1608} = .86.$$

These general observations will be of value in Section IV.

Example 3.

In Table X the S-Array and Columns 1 and 4 of the R-Array are displayed for the process

$$X_t - .8X_{t-1} - .86X_{t-2} + .83X_{t-3} = Z_t - .9Z_{t-1}. \qquad (18)$$

They were calculated using $\hat{\rho}_k$ which was obtained from 300 simulated values from (18) with Z_t i.i.d. N(0, 1).

TABLE X

R & S ARRAYS at $f_m = (-1)^m \hat{\rho}_m$ for Data Generated from (18)

$S_1(m)$	$S_2(m)$	$S_3(m)$	$S_4(m)$	m	$R_1(m)$	...	$R_4(m)$
-.3166	-.0730	.1511	-.0843	-9	.7040	...	.3956
.4645	-.0196	.1389	-.1621	-8	.4810	...	.1556
-.1948	-.1069	.1537	-.5727	-7	.7045		-.0085
.1833	.2332	.1532	-.1693	-6	.5672		.0234
.0580	-.0120	.1530	90.8817	-5	.6712		-.0024
-.1301	-.1479	.1526	.3836*	-4	.7102		-1.4827
.3996	-.0554	.2829*	.0794*	-3	.6178		-3.4936*
-.3326	-.1075*	-.0658*	.1375	-2	.8647		.7232
.7327*	.0857*	-.1379	1.4015	-1	.5771	...	.0022
-.4228*	.0407	-.1377	.1837	0	1.0000*		-.0205
.4984	.0879	-.1346	2.4277	1	.5771		.0075
-.2855	.0046	-.1345	-.1535	2	.8647		-.1279
.1496	.1240	-.1264	-.0368	3	.6178		-.2833
-.0548	.3387	-.0994	-.0157	4	.7102		-.3881
-.1549	.0291	-.1115	.0334	5	.6712		-.4429

Again it is very clear from the S-ARRAY that p = 3 and q = 1. Note the behavior of both Columns 3 and 4 are exactly as would be expected for a ARMA (3, 1). That is, Column 3 has the characteristic "constant behavior", Column 4 is highly variable with $S_4(-2) \approx S_3(-1)$ while $|S_4(-5)|$ is by far the biggest number in the table. Column 4

of the R-array also clearly gives $q = 1$. Thus again it is clear that no other values of p and q seem even remotely possible.

Example 4. (Sunspot Data)

In this example we consider the sunspot data considered in Box and Jenkins and denoted as Series E. The data consists of 100 data points referred to as sunspot numbers and can be found in [5]. Table XI gives the sample acf for this series while Table XII shows the corresponding S-Array at $f_m = (-1)^m \hat{\rho}_m$. The R-array is not given here since the S-array is so distinctive. Both the behavior of the second and third column are distinctive here and clearly we have $p = 2$, $q = 0$ which is the same as found in [5]. It is worth noting here that Column 2 is surprisingly stable for such a short series as this. Possible explanations for this behavior are hidden periodicities and nonstationarity, and are currently being considered by the authors.

TABLE XI

$\hat{\rho}_m$ for Sunspot Data

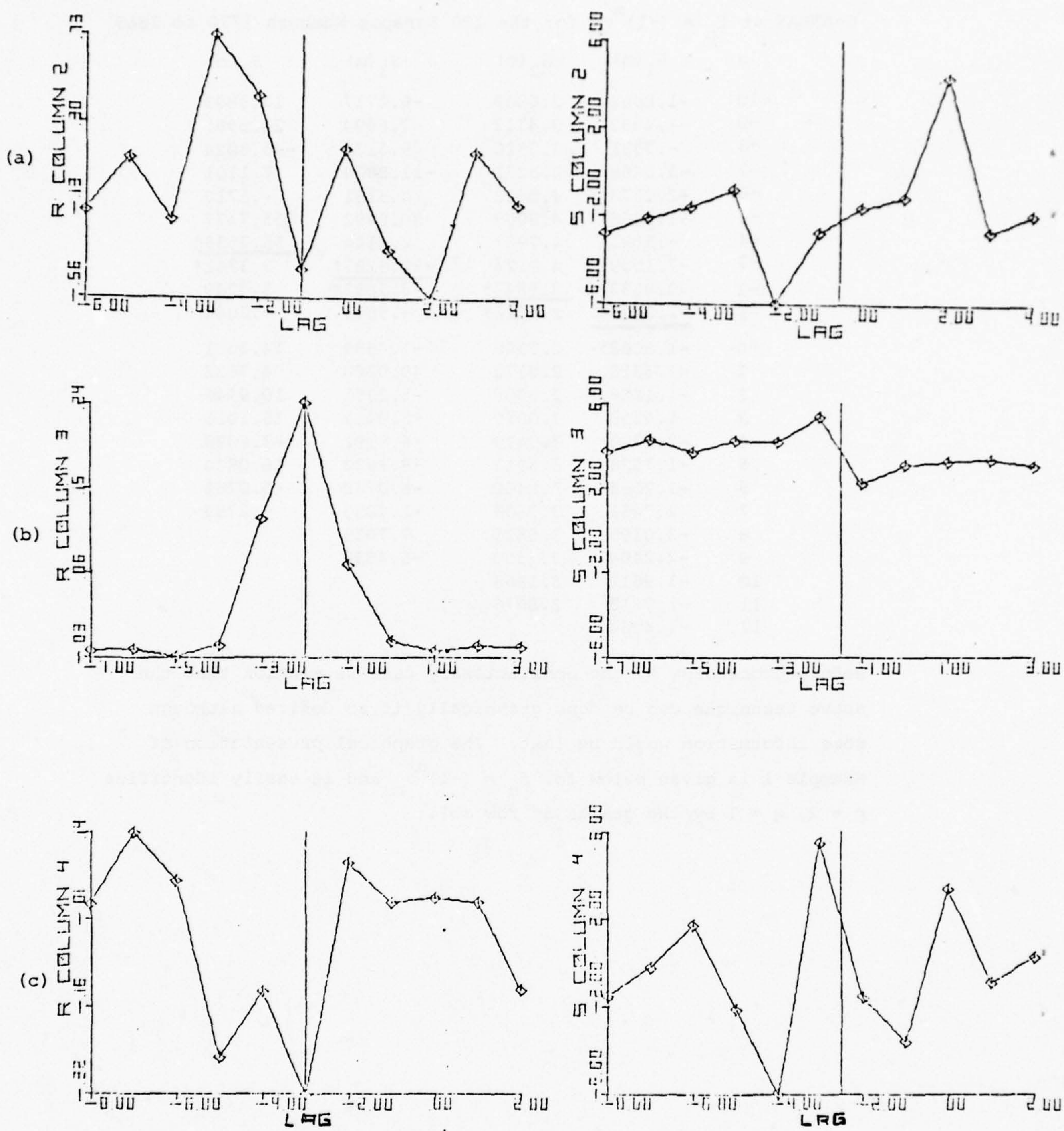
m							
1-7	1	.806	.428	.070	-.169	.-266	-.212
8-14	-.044	.169	.330	.410	.394	.288	.143

TABLE XII

S-ARRAY at $f_m = (-1)^m \hat{\rho}_m$ for the 100 Sunspot Numbers 1770 to 1869

m	$S_1(m)$	$S_2(m)$	$S_3(m)$	$S_4(m)$
-10	-1.8062	3.6869	-6.2717	16.9809
-9	-1.4952	3.4112	-7.6894	24.5965
-8	-.7331	3.3526	-9.3130	-45.6024
-7	-5.8468	3.5271	-11.2804	7.1101
-6	-2.2570	4.0622	4.5181	-.1710
-5	-1.6366	4.8009	-30.8892	55.7477
-4	-.5891	4.7937	2.4144	55.3535*
-3	-7.1500	4.9124	-39.6295*	3.3841*
-2	-2.8833	4.6547*	-3.1892*	3.3249
-1	-2.2403*	2.9516*	-.9613	-.2644
0	-1.8062*	2.8146	-3.4689	14.3601
1	-1.5310	2.9172	48.0260	4.7632
2	-1.1626	2.9305	-5.2351	10.9746
3	1.4339	3.0070	-5.9414	15.1016
4	-2.5709	3.3499	-6.8192	-3.6098
5	-1.7956	3.6912	-8.2632	16.0930
6	-1.2063	3.8460	-6.0742	-5.0782
7	2.7461	3.7688	-1.7353	-.2789
8	-3.0195	3.5825	7.7039	
9	-2.2404	3.7350	-5.4833	
10	-1.9612	3.1968		
11	-1.7313	2.8076		
12	-1.4964			

Before proceeding to the nonstationary case we mention that the above technique can be done graphically if so desired although some information would be lost. The graphical presentation of Example 1 is given below for $f_m = (-1)^m \hat{\rho}_m$ and it easily identifies $p = 2$, $q = 1$ by the graphs of row (b).



Graphical Presentation of R and S Arrays

IV. THE D(n,m) STATISTIC

The identification procedure thus far suggested is one of pattern recognition. In this section a statistic is suggested to measure agreement with the proper pattern for a stationary ARMA (p, q) process. There is no question that some information will be lost in this statistic. However, as will be seen, it is sufficiently sensitive to different values of p and q that in many problems it will be quite adequate for identifying p and q. In fact in all of the previous examples this statistic, which we refer to as the D-Statistic, easily identifies the "proper" values of p and q. In some later small sample examples the statistic shows appropriate contending models and suggests the obvious fact that a visual examination of the S and R arrays is in order.

Definition 4.

For each pair of integers (n,m) such that n = 0, 1, 2, ..., N and m = 0, 1, 2, ..., M, define

$$D(n,m) = \frac{\left|\dfrac{S_{n+1}(f_{-m-n-1})}{S_n(f_{-m-n})}\right|}{V},$$

$$V = \left[\frac{S_n(f_{m-n+1}) + S_{n+1}(f_{m-n})}{S_{n-1}(f_{m-n+2}) + S_n(f_{m-n+1})}\right]^2 + \sum_{i=1}^{3}\left[\left(\frac{r_{n+1}(f_{-m-n-i})}{r_n(f_{-m-n-i+1})}\right)^2 + \left(\frac{r_{n+1}(f_{m-n+i})}{r_n(f_{m-n+i+1})}\right)^2\right],$$

where $S_{-1}(f_m) \equiv 0$ and $r_o(f_m) \equiv 1$.

Inspection of Tables III and IV as well as the recursion relations (4) and (5) furnish the motivation for our definition of D(n,m). Although not clear from Tables III and IV, the relations (4) and (5) show the zero behavior in the R-array will tend to continue in columns

p+2, p+3, ... due to the multiplication by r_n in the calculation of r_{n+1}. Also from those relations it is clear the "infinity and $-c_1$ effect" will tend to continue in columns p+1, p+2, ... This effect is nullified however in D(n,m) by the divisions in the numerator and denominator and hence a large value of D(n,m) is suggestive of agreement with an ARMA (n,m) process. Thus the D-estimate for p and q is defined by the rule, select the model ARMA (p,q) where

$$D(p,q) = \max_{\substack{n=1,\ldots,N \\ m=0,\ldots,M}} D(n,m).$$

The case n = 0 is excluded in the definition of D(p,q) since it is not "normalized". See Sections VII and IX.

Although these authors feel it is always instructive to inspect the R and S arrays, when D(p,q) is as distinctive as it is for the examples just considered, the statistic will normally suffice if stationarity is assured.

However if the process is nonstationary or seasonal in the sense that columns in the R-array have runs of "near zeros" interlaced among values not near zero, the D-Statistic would not normally be of value and hence as the remaining part of the paper will show a visual examination of the R and S arrays is usually better. We now list a short list of D(n,m) values for the previous examples. In each case the correct value of p and q is clearly indicated.

D(n,m)

(n,m)	Example # 1	2	3	4
(1,0)	.013	16,603.009	.004	.016
(1,1)	.001	.175	.004	.012
(1,2)	1.041	.107	.011	.003
(1,3)	1.045	.306	.004	.675
(2,0)	.811	.001	.021	170.680
(2,1)	110.755	.012	.004	.232
(2,2)	.590	.001	.001	38.208
(2,3)	.042	.001	.018	.007
(3,0)	.003	.004	.566	.002
(3,1)	.001	.036	1,313,892.061	.087
(3,2)	.035	.088	.001	.000
(3,3)	.004	.000	15.241	.489

In Section VII we will investigate smaller sample cases and D(n,m) will not be nearly so distinctive as these examples. Nevertheless it will be found to be useful even in small sample cases.

A limited Monte Carlo study comprised of approximately 10 different processes and 50 repetitions of each at a variety of sample sizes was performed by the authors to compare the D-Statistic as an automated method for estimating p and q with existing methods. The methods compared against were the FPE and AIC methods suggested by Akaike (1,2) and a method suggested by Anderson (3). Of course the FPE and Anderson method must always select an AR(p) process but the comparisons were still of interest. For strictly autoregressive processes the Anderson method always did the best with the D-Statistic close behind. For processes in general, i.e., ARMA (p, q), the D-Statistic performed much better than its only competitor, the AIC criteria. In addition to correctly identifying p,q far more often than the AIC criteria, the computation time for the D-Statistic is a great deal smaller.

V. ARMA (p, q) PROCESSES - THE NONSTATIONARY CASE

In our previous examples we have assumed a stationary model. For most problems this is not a realistic assumption and some attempt must be made to transform the data to stationary data. In the case of ARMA (p, q) modeling the most common approach is to attempt to fit an ARIMA (p, d, q) model. That is, observe the sample autocorrelation, if it fails to damp out "sufficiently fast", difference the data (possible several times) until the resulting data produces an autocorrelation which damps out rapidly. Of course even though the process may be ARMA (p, q) with roots on the unit circle, differencing may not produce a stationary process. In fact it is clear that for such models the procedure will only be beneficial when 1 is "almost a solution" of the characteristic equation.

In this section the theorems of the previous sections for model identification will be generalized to the nonstationary ARMA (p, q) process. As a result it will be shown how a transformation to stationarity may be obtained anytime the process is nonstationary due to roots, real or complex, on the unit circle. Recall that an ARMA (p, q) process is nonstationary if and only if the roots of the characteristic equation

$$1 - \phi_1 x - \phi_2 x^2 - \ldots - \phi_p x^p = 0 \tag{19}$$

lie outside the unit circle. If the roots lie much in the interior of the circle then the time series is quite explosive and easily recognized as nonstationary. In this paper we will not consider such series but will concentrate on nonstationarities due to roots on the unit circle. Thus in the remaining portion of the paper our use of the word nonstationary will mean that the characteristic equation has roots <u>on</u> the unit circle. We give the following definition and theorem for future reference.

<u>Definition 5</u>.

An ARMA (p, q) process will be said to be purely nonstationary if all of the roots of its characteristic equation lie on the unit circle.

<u>Theorem 6</u>.

Let $r_1, r_2, \ldots, r_p$ be the roots of the characteristic equation $\phi(x) = 1 - \phi_1 x - \phi_2 x^2 - \ldots - \phi_p x^p = 0$ of an ARMA (p, q) process and suppose the ϕ_i's are such that $|r_i| \geq 1$ for $i = 1, 2, \ldots, p$. Then $|\phi_p| = 1$ iff the process is <u>purely</u> nonstationary.

<u>Proof</u>. Since

$$\phi_p = (-1)^{p+1} \prod_{i=1}^{p} r_i^{-1}$$

we have

$$|\phi_p| = |r_1^{-1}| \ldots |r_p^{-1}| .$$

But since $|r_i| \geq 1$ for all i the theorem follows at once.

In order to interpret ρ_m when the underlying process is nonstationary we make the following definition.

Definition 6.

Denote by $\rho_m(\alpha_1, \alpha_2, \ldots \alpha_p)$ the autocorrelation of an ARMA (p, q) process with roots of the characteristic equation α_i, such that $|\alpha_i| > 1$. Let $r_1, r_2, \ldots r_p$ be the roots of the characteristic equation of a given ARMA (p, q) process with j roots on the unit circle, which for convenience we denote by $r_1, r_2, \ldots r_j$. Then define

$$\rho_m^* = \rho_m(r_1, r_2, \ldots r_p) \quad , \quad j = 0$$

and

$$\rho_m^* = \lim_{|\alpha_1|,\ldots|\alpha_j|\to 1^+} \rho_m(\alpha_1,\alpha_2,\ldots, \alpha_j,r_{j+1}\ldots r_p) \quad , \ j = 1,2\ldots p.$$

where $|\alpha_1| = |\alpha_2| = \ldots |\alpha_j|$.

The condition in Definition 6 restricting the path along which the limit is taken is necessary to guarantee the existence of the limit. It is compatible with the idea that in practice roots "near" and equally close to unit circle would be classified as on the unit circle. Henceforth we let $\rho_m = \rho_m^*$ so that ρ_m is well defined for both stationary and nonstationary processes.

We are now in a position to show the basic result which we will utilize to detect nonstationarity in data.

Theorem 7.

Let $\rho_m = \rho_m^*$ and assume $|r_i| \geq 1$. Suppose there exists an $\ell > 0$ which is the smallest integer such that $S_\ell(f_m) = C_f$ for $m = 0 \mp 1, \mp 2, \ldots$, where C_f is a constant depending on whether $f_m = \rho_m$ or $f_m = (-1)^m\rho_m$. Then the process is nonstationary.

Proof. Assume w.l.o.g. that $f_m = \rho_m$. Then if $S_\ell(m) = C_f \neq 0$ for $m = 0 \mp 1, \mp 2, \ldots$, it follows from Theorem 8 of [7] that ρ_m satisfies an ℓ^{th} order difference equation. More specifically, there exists constants $\phi_1, \phi_2, \ldots, \phi_\ell$ such that

$$\rho_m = \phi_1\rho_{m-1} + \ldots + \phi_\ell\rho_{m-\ell} \ , \quad \text{for } m = 0, \mp 1, \mp 2, \ldots .$$

But $S_\ell(m) \equiv C_f \neq 0$ implies from the proofs of Theorems 2 and 3 that $\phi_\ell = -1$ and from Theorem 6 ρ_m satisfies the characteristic

equation of a purely nonstationary process and hence $j \geq 1$ in Definition 6 and the process is nonstationary. If $C_f = 0$, since ℓ is minimal, $H_{\ell+1}(1;k)_\ell = 0$. However the proof of Theorem 8 in [7] then implies $1 - \sum_{i=1} \phi_i = 0$ which again implies $j \geq 1$ so that the process is nonstationary.

In order to exemplify Theorem 7, the S-Array at $f_m = (-1)^m \hat{\rho}_m$ for a realization of length 200 from the AR(2) process

$$(1-.99B)(1-.5B)X_t = X_t - 1.49X_{t-1} + .495X_{t-2} = Z_t, \quad (20)$$

where $Z_t \sim N(0, 1)$, followed by the S-Array at $f_m = (-1)^m \rho_m$, is given in Table XIII. In this case the root of interest (call it r_1) is slightly outside the unit circle. If it were exactly 1, as a later result will show, the S_1 column would be identically -2. Thus in that event $\ell = 1$ in Theorem 7. The process of Equation (20) is of course mathematically stationary. However, in practice such an equation is a nonperfect model and consequently for most purposes we would prefer to classify the model as nonstationary when it has an estimated root so near the unit circle. Thus if such a model were obtained via a set of data the factor (1-.99B) would be altered to (1-B). This will in general be our modus operandi for roots indicated near the unit circle, real or complex. Moreover since our use of the S-Array will therefore be to detect an indication of a root near the unit circle our simulated examples will be for "nearly nonstationary" processes. It should be noticed that when the sample autocorrelation is utilized in Table XIII that the constant behavior which is supposed to appear in column two if the process is stationary has been completely obscured by the nonstationary nature of the data. The next few theorems and examples should make clear this behavior as well as establish how to deal with it. Before leaving this example however let us note that the autocorrelation here tends to a solution of a first order difference equation as $r_1 \to 1$. Thus the $\lim_{r_1 \to 1} \rho_m(r_1, r_2)$ satisfies a lower order difference equation that the characteristic equation, namely the one generated by the nonstationary factor.

That is in this case if $\rho_m^* = \lim_{r_1 \to 1} \rho_m(r_1, r_2)$ we have $(1-B)\rho_m^* = 0$, $m = 0, \mp 1, \ldots$, and it is clear why in Table XIII, $S_1(m) \approx -2$. Since in practice one essentially assumes the autocorrelation of a nonstationary process is a limit such as ρ_m^*, Definition 6 and Theorem 7 are of practical importance.

TABLE XIII

$f_m = (-1)^m \hat{\rho}_m$

m	$S_1(m)$	$S_2(m)$	$S_3(m)$
-8	-2.032	1.826	-3.438
-7	-2.034	-5.808	-5.032
-6	-2.034	4.214	-52.454
-5	-2.034	2.895	-21.068
-4	-2.035	-33.755	11.327
-3	-2.035	5.702	37.676*
-2	-2.031	16.718*	-2.115*
-1	-2.024*	2.240*	-1.233
0	-1.976*	3.053	-2.542
1	-1.969	1.854	-4.025
2	-1.965	6.615	-8.957
3	-1.966	3.803	6.454
4	-1.966	1.456	5.604
5	-1.967	-17.359	-12.754
6	-1.966	-1.906	-5.647
7	-1.968	4.721	-2.597
8	-1.965	2.426	-20.306

$f_m = (-1)^m \rho_m$

m	$S_1(m)$	$S_2(m)$	$S_3(m)$
-8	-2.010	6.030	u
-7	-2.010	6.030	u
-6	-2.010	6.030	u
-5	-2.010	6.030	u
-4	-2.009	6.030	u
-3	-2.008	6.030	∞*
-2	-2.007	6.030*	-2.985*
-1	-2.003*	2.985*	u
0	-1.997*	2.985	u
1	-1.993	2.985	u
2	-1.992	2.985	u
3	-1.991	2.985	u
4	-1.990	2.985	u
5	-1.990	2.985	u
6	-1.990	2.985	u
7	-1.990	2.985	u
8	-1.990	2.985	u

Theorem 8.

Suppose an ARMA (p, q) process has k roots of the characteristic equation which approach the unit circle as in Definition 6, N of which are distinct. Then for $m = 0, \mp 1, \ldots, \rho^*_m$ satisfies a linear homogeneous difference equation of order less than or equal to N, whose characteristic equation has all of its roots distinct and on the unit circle.

Proof. The proof of Theorem 8 is included in Appendix 3 for an ARMA (3, 1) process only. It was first shown by Kelley in [8] where the cases ARMA (1, 0), (1, 1), (1, 2), (2, 0), (2, 1), (2, 2) and (3, 0) are also given. In addition the result is demonstrated for an ARMA (4, 1) in Example 6. An analytic proof of higher order cases seems to be intractable, at least by Kelley's approach, due to the lengthy algebra involved. However since Theorem 8 and Theorem 9 which follow will only be utilized to suggest an appropriate transformation to stationarity, we shall be satisfied with the validity of the result as given. That is, once the transformation is constructed, it can be observed whether or not it had the desired effect. As our final theorem we make us of Theorem 8 to obtain the converse of Theorem 7.

Theorem 9.

Suppose Theorem 8 holds then the condition that $S_\ell(m)$ is constant for some $\ell \leq p$ is both necessary and sufficient for nonstationarity.

Proof. The sufficiency follows of course from Theorem 7. Now let $f_m = \rho_m$ and suppose the process is nonstationary with autocorrelation $\rho_m = \rho^*_m$. Then let N be the number of distinct roots approaching the circle in the manner described by Definition 6. By Theorem 8, ρ_m satisfies the characteristic equation of a purely nonstationary AR process of order $\ell \leq N$, i.e.

$$(1 - \psi_1 B - \psi_2 B^2 - \ldots - \psi_\ell B^\ell)\rho_m = 0 \qquad (21)$$

and

$$|\psi_\ell| = 1 \ .$$

But since ρ_m satisfies an equation such as (21), if $1 - \sum_{1}^{\ell} \psi_\ell \neq 0$, the proof of Theorem 1 (see [6,7]) implies Theorem 2 and Theorem 3 hold even though the process is nonstationary.

Thus $\psi_\ell = -1$ implies $C_1 = C_2$ or $S_\ell(k)$ constant. On the other hand if $\psi_\ell = 1$, since all the roots are on the unit circle, it is easily shown that $1 - \sum_{i=1}^{\ell} \psi_i = 0$. However in this case the proof given of Theorem 1 in [6] can be modified to show $S_\ell(m) \equiv 0$ so that again we have $S_\ell(m)$ constant. The proof for $f_m = (-1)^m \rho_m$ is entirely similar and hence is not argued. Before proceeding we list three useful corollaries to the above results whose proofs follow from the proofs of those results.

<u>Corollary 2</u>. If $S_1(m) \equiv 0$ when $f_m = \rho_m$ or if $S_1(m) \equiv -2$ when $f_m = (-1)^m \rho_m$, then the process is nonstationary and the characteristic equation has at least one root of +1 on the unit circle.

<u>Corollary 3</u>. If $S_1(m) \equiv -2$ when $f_m = \rho_m$ or $S_1(m) \equiv 0$ when $f_m = (-1)^m \rho_m$, then the process is nonstationary and the characteristic equation has at least one root of -1.

<u>Corollary 4</u>. If $S_2(m) \equiv 0$ for $f_m = \rho_m$ or $f_m = (-1)^m \rho_m$, then the process is nonstationary and the characteristic equation has at least one root of +1 and at least one root of -1. A number of results such as Corollaries 2, 3 and 4 could be listed. They of course demonstrate the manner in which evidence from the S-Array can be useful in determining the appropriate transformation to stationarity. We now list a final corollary of this nature. Others will become apparent in the next section.

Corollary 5. If $S_2(m) \equiv C \neq 0$, then the process is nonstationary and the characteristic equation has at least two complex roots on the unit circle. Tables XIV and XV illustrate the above results as a root approaches 1^+ for an ARMA (1, 1) and an ARMA (3, 1). In both cases the convergence of $S_1(k)$ to -2 is clear even though the sample autocorrelations were utilized in the calculations. Example 6 which follows demonstrates Corollary 5. Theorem 7 is demonstrated graphically in Figures 1, 2 and 3. It should be noted that although the autocorrelations are dramatically different, the $S_\ell(m)$ columns are essentially the same. That is, they are roughly constant with possibly some slight deviation about the "center line", $k = -\ell + \frac{1}{2}$. The reason for the possible deviation can be seen from Tables XIV and XV where it is clear that the elements $S_1(-1)$ and $S_1(0)$, due to the moving average, behave somewhat differently than the remaining elements although that difference, as it must, is dimensioning as $G_1 \to 1$.

Let us now consider some additional examples which make use of all of our results to this point.

VI. APPLICATIONS

The examples which follow include some from generated data and some from real data. In the case of generated data, the data (as in the previous examples) was generated on the CDC Cyber 72 via the ARMA (p, q) equation with Gaussian white noise input unless stated to the contrary.

Example 5. (Series C)

The source of data for this example are 226 temperature readings from a chemical process taken every minute. The data is given in [5] where it is referred to as Series C. The data is plotted below in Figure 4.

Fig. 1 Based on 250 points from $(1-.98B)(1+.6B)X_t = (1+B+.02B^2-.232B^3)Z_t$

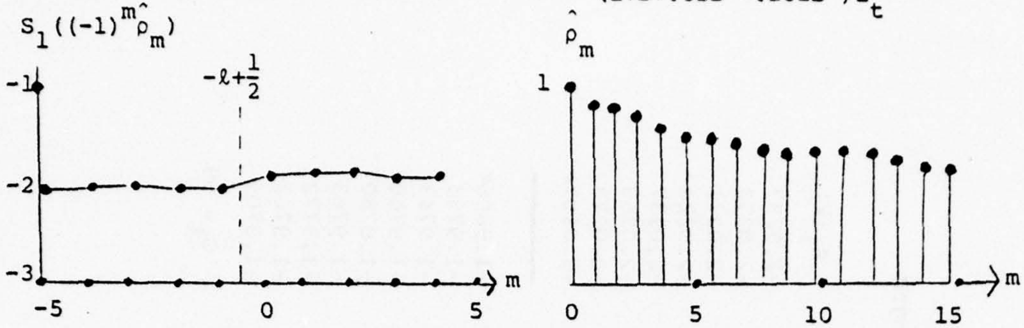


Fig. 2 Based on 250 points from $(1-.99B)(1+.99B)(1-.66B)X_t = (1-.4B+.2B^2)Z_t$

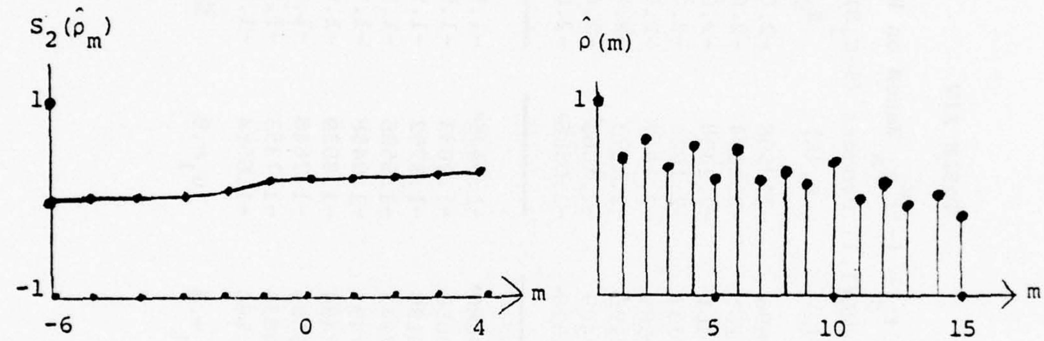


Fig. 3 Based on 300 points from $\prod_{i=1}^{4}(1-G_iB)X_t = Z_t$

$G_1 = .98,\ G_2 = .2 + .96i,\ G_3 = .2 - .96_i,\ G_4 = -.4$

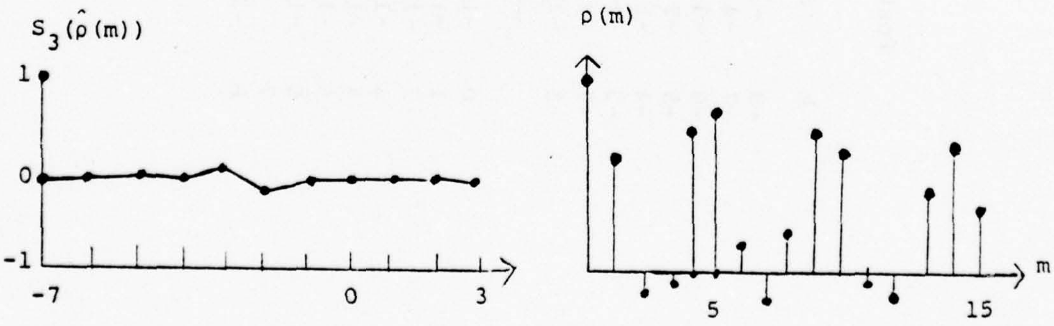


TABLE XIV

Portions of $\underline{S}_1$ at $f_m = (-1)^m\hat{\rho}_m$ Based on Realizations of Length 200 of the ARMA(1,1) Process $(1-G_1B)X_t = (1+.35B)Z_t$

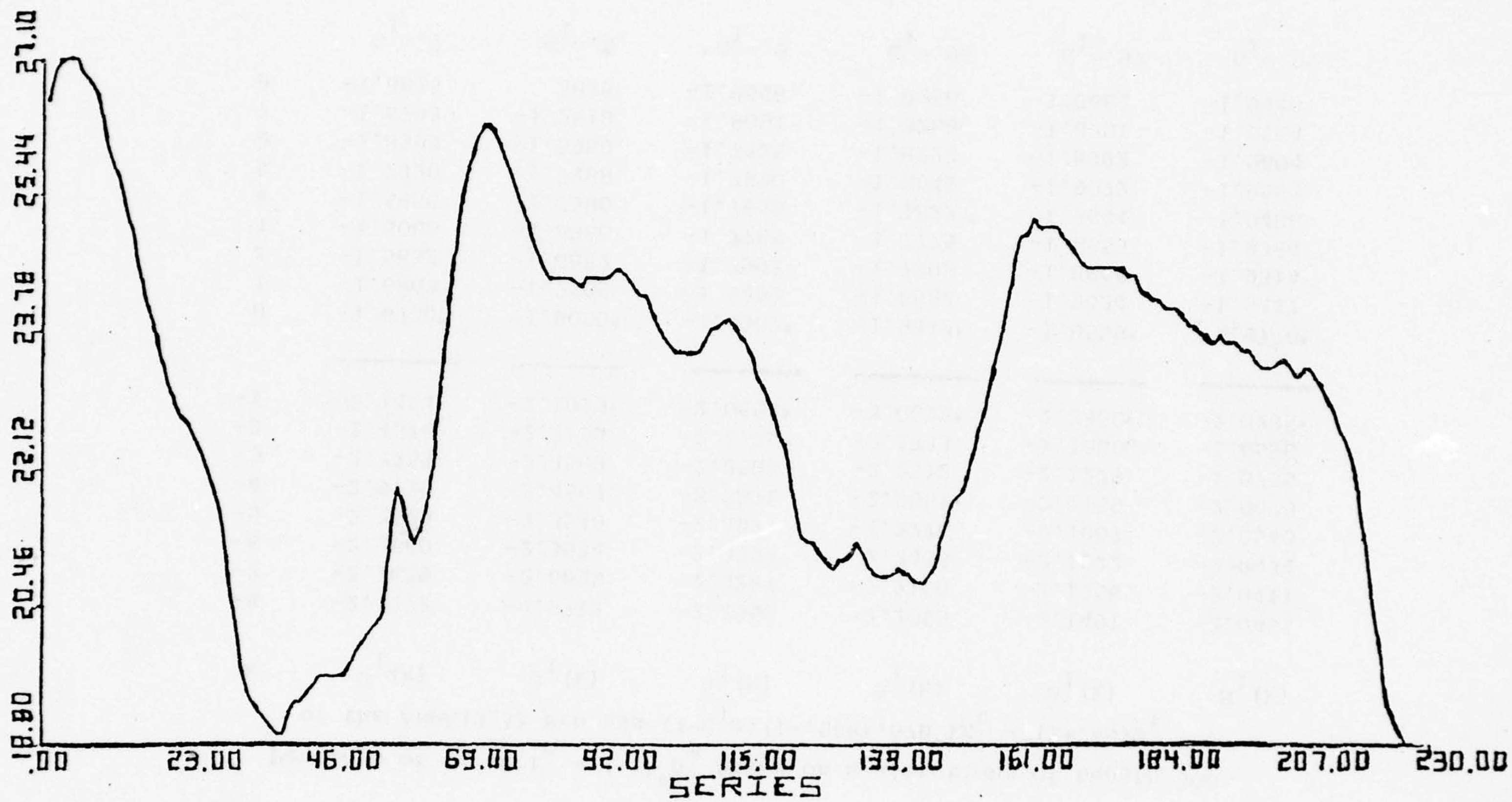
k	$S_1(k)$	$S_1(k)$	$S_1(k)$	$S_1(k)$	$S_1(k)$	$S_1(k)$
-8	-.3474	-3.0696	-2.3598	-2.0811	-2.0633	-2.0237
-7	-4.6339	-2.5181	-2.3074	-2.0747	-2.0541	-2.0233
-6	-3.3134	-2.3361	-2.2438	-2.0787	-2.0615	-2.0221
-5	-2.9187	-2.3337	-2.1823	-2.0704	-2.0679	-2.0224
-4	-2.3673	-2.2589	-2.1647	-2.0668	-2.0680	-2.0237
-3	-2.4190	-2.2196	-2.1373	-2.0611	-2.0615	-2.0263
-2	-2.4988	-2.2409	-2.1069	-2.0548	-2.0661	-2.0273
-1	-2.2499*	-2.1520*	-2.0586*	-2.0348*	-2.0500*	-2.0197*
0	-1.8000*	-1.8680*	-1.9445*	-1.9663*	-1.9523*	-1.9806*
1	-1.6671	-1.8058	-1.9033	-1.9479	-1.9379	-1.9733
2	-1.7046	-1.8199	-1.8792	-1.9423	-1.9420	-1.9743
3	-1.7313	-1.7943	-1.8585	-1.9373	-1.9362	-1.9768
4	-1.5211	-1.7497	-1.8458	-1.9342	-1.9363	-1.9780
5	-1.4322	-1.7484	-1.8039	-1.9270	-1.9420	-1.9783
6	-1.2751	-1.6586	-1.7648	-1.9304	-1.9486	-1.9772
7	.5324	-1.4831	-1.7353	-1.9249	-1.9404	-1.9767
8	-3.2324	-1.2180	-1.6564	-1.9219	-1.9389	-1.9709
	$G_1=.7$	$G_1=.8$	$G_1=.9$	$G_1=.95$	$G_1=.97$	$G_1=.99$

TABLE XV

Portions of $\underset{\sim}{S}_1$ at $f_m = (-1)^m \hat{\rho}_m$ Based on Realizations of Length 300 of the ARMA(3,1) Process $(1-G_1B)(1-.90B+.67B^2)X_t = (1+.4B)Z_t$

k	$S_1(k)$	$S_1(k)$	$S_1(k)$	$S_1(k)$	$S_1(k)$	$S_1(k)$
-8	-2.1372	-4.9712	-2.2358	-2.2065	-2.1491	-2.0455
-7	-2.1629	-2.6695	-2.3287	-2.2160	-2.1258	-2.0411
-6	-2.3660	-2.3534	-2.3297	-2.2436	-2.1022	-2.0434
-5	-2.7679	-2.3548	-2.3071	-2.2776	-2.1007	-2.0540
-4	-2.9730	-2.4563	-2.2871	-2.2861	-2.1169	-2.0679
-3	-2.7753	-2.4592	-2.2509	-2.2512	-2.1279	-2.0735
-2	-2.4570	-2.3158	-2.1775	-2.1733	-2.1080	-2.0596
-1	-2.1454*	-2.1099*	-2.0652*	-2.0622*	-2.0460*	-2.0235*
0	-1.8730*	-1.9009*	-1.9387*	-1.9414*	-1.9559*	-1.9770*
1	-1.6863	-1.7599	-1.8492	-1.8522	-1.9024	-1.9437
2	-1.5632	-1.6852	-1.7993	-1.7992	-1.8865	-1.9314
3	-1.5068	-1.6866	-1.7769	-1.7775	-1.8953	-1.9364
4	-1.5656	-1.7380	-1.7650	-1.7827	-1.9084	-1.9486
5	-1.7320	-1.7388	-1.7520	-1.8041	-1.9072	-1.9583
6	-1.8599	-1.5989	-1.7525	-1.8223	-1.8882	-1.9604
7	-1.8793	-1.2518	-1.8091	-1.8288	-1.8701	-1.9564
8	-1.8825	.5835	-1.9558	-1.8250	-1.8683	-1.9485
	$G_1=.7$	$G_1=.8$	$G_1=.9$	$G_1=.95$	$G_1=.97$	$G_1=.99$

Fig. 4 A Graph of Series C



Tables XVI and XVII show the S-array for the data. Table XVI clearly identifies the process as essentially nonstationary due to a unit root. In addition that array seems to clearly indicate the process is an ARMA (2, 1). However, a word of caution is in order. That is, as we have previously seen and explained, the apparent nonstationarity will often completely dominate the column behavior. In Table XVI it is not clear that this is the case and the ARMA (2, 1) shows through remarkedly well. Nevertheless the non-stationarity should be removed before the final identification. Table XVII shows the S-array for the first differences of Series C. The differences are by that table a stationary AR (1) with

$$\hat{\phi} = \frac{1.805}{2.241} \approx .81 \ .$$

Thus the model obtained is

$$(1-B)(1-.81B)\ X_t = Z_t \ . \tag{22}$$

It should now be clear that the apparent moving average term indicated in Table XVI is indeed induced by the nonstationarity which is vividly displayed in Column 1 of that table. As previously explained this is to be expected since ρ_m^* corresponding to (22) actually satisfies a first order DE. The remarkable thing here is that the 2nd order process shows up at all in Table XVI. Before leaving this example let us make a point which will be useful in the next example. That is, notice that we have actually obtained estimates of the coefficients one factor at a time. From $S_1(m)$ of Table XVI we obtained the factor (1-B) and from Table XVII we obtained the estimate $\hat{\phi}$. The resulting model is the same as the one given for this data in Box and Jenkins [5]. The next example shows how this observation can be extended.

Example 6.

We now consider an example of an ARMA (4, 1) process with two complex roots nearly on the unit circle. The data consists of 300 points generated from the process

$$(1-1.2B+.45B^2)(1-1.7B+.99B^2)\ X_t = (1-.5B)\ Z_t \tag{23}$$

TABLE XVI

A Portion of the S-Array at

$f_m = (-1)^m \hat{\rho}_m$

for the Series C Data

m	S_1	S_2	S_3
-7	-2.080	4.755	-2.464
-6	-2.073	4.314	-5.288
-5	-2.064	4.471	5.590
-4	-2.056	4.497	567.959
-3	-2.046	4.449	13.380*
-2	-2.035	9.576*	-2.100*
-1	-2.022*	2.492*	-3.544
0	-1.977*	3.604	-1.989
1	-1.965	3.572	-23.995
2	-1.955	3.589	4.549
3	-1.946	3.704	2.909
4	-1.939	3.413	-1.719
5	-1.931	4.060	-4.178
6	-1.925	3.660	31.451
7	-1.918	3.604	-2.994

TABLE XVII

A Portion of the S-Array at

$f_m = (-1)^m \hat{\rho}_m$

for the First Difference of the Series C Data

m	S_1	S_2
-7	-2.217	5.250
-6	-2.192	-.212
-5	-2.163	7.000
-4	-2.190	1.911
-3	-2.240	-.969
-2	-2.234	-170.077*
-1	-2.241*	1.786*
0	-1.805*	.546
1	-1.810	-10.704
2	-1.806	2.706
3	-1.840	.164
4	-1.859	3.127
5	-1.838	11.435
6	-1.821	1.314
7	-1.712	12.830

A graph of the data is given in Figure 5 while a graph of the autocorrelations is given in Figure 6. In Table XVIII the S-Array for this data is shown out to Column 5. The fact that $S_2(m)$ is nearly a nonzero constant immediately implies the process has two complex roots very near the unit circle and that the sample autocorrelation behaves as the autocorrelation of a purely nonstationary AR (2) process with $\psi_2 \sim -1$. Thus we assume the process to have the factor $1 - \psi_1 B + B^2$. But since the sample acf behaves as that of an AR (2) we have from our previous results that

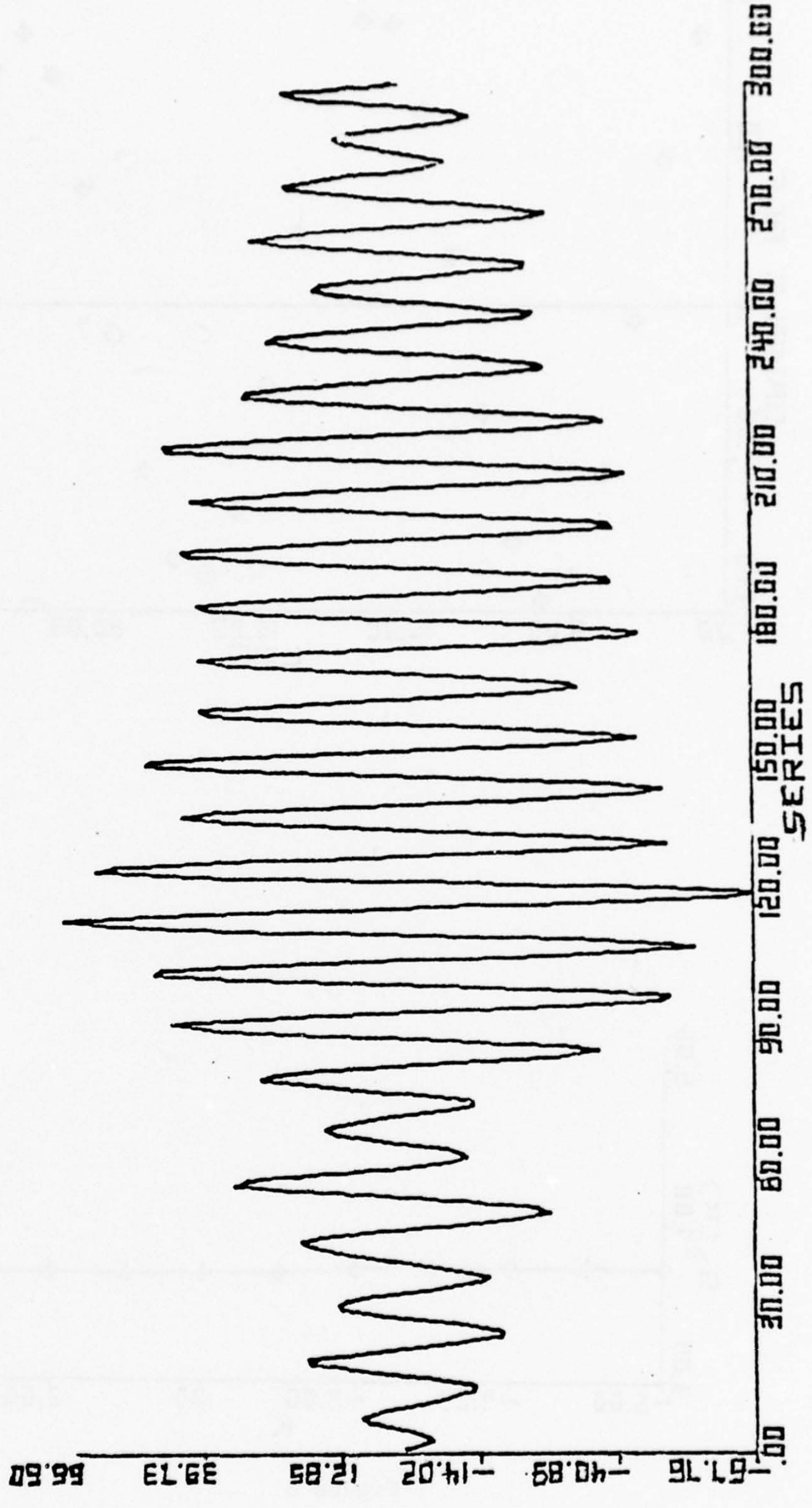
$$1 + \psi_1 - \psi_2 = C$$

or

$$\psi_1 = C - 2$$

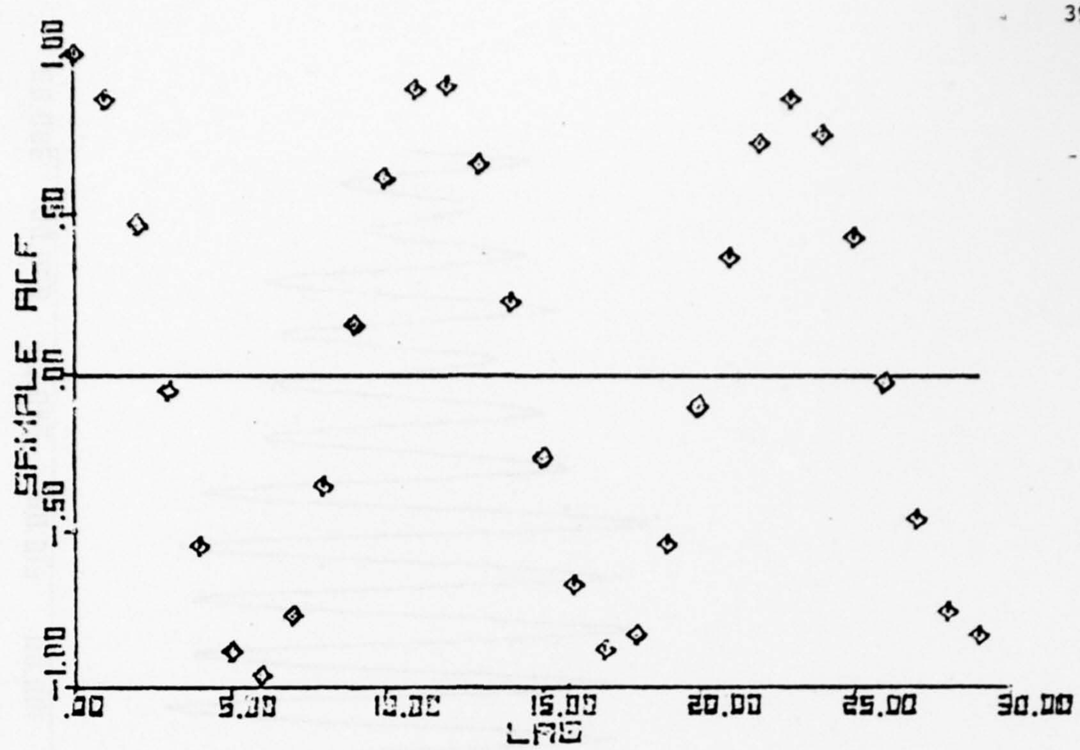
Now if we estimate C by averaging several of the elements in Column 2 surrounding the starred values, say $S_2(-4)$ to $S_2(1)$, we obtain the estimate $\hat{C} = 3.71$ and hence

$$\hat{\psi}_1 = 1.71$$



-67.76
-40.89
-14.02
12.85
39.73
66.60
.00
30.00
60.00
90.00
120.00
150.00
180.00
210.00
240.00
270.00
300.00
SERIES

Figure 5


SAMPLE ACF
1.00
.50
.00
-.50
-1.00
.00
5.00
10.00
15.00
20.00
25.00
30.00
LAG

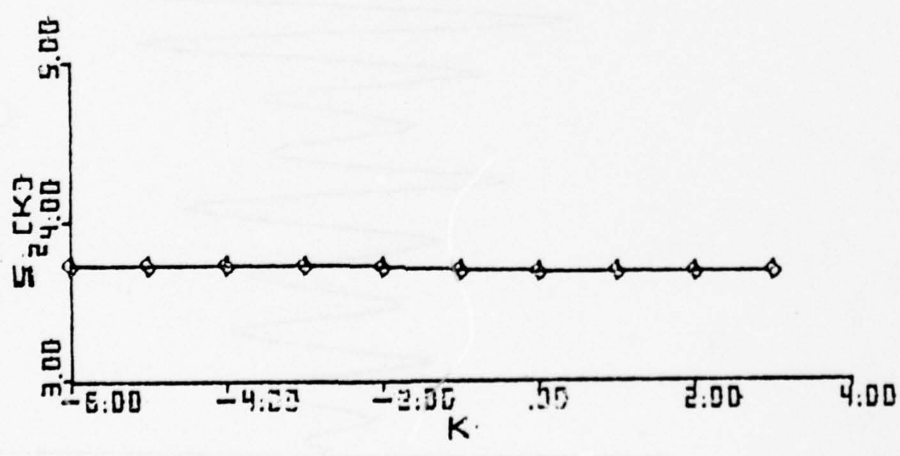

S2(K)
5.00
4.00
3.00
-6.00
-4.00
-2.00
.00
2.00
4.00
K

Figure 6

TABLE XVIII

A Portion of the S-Array at $f_m = (-1)^m \hat{\rho}_m$ for a Realization of Length 300 of the ARMA(4,1) Process in Example 6

k	$\underline{S}_1$	$\underline{S}_2$	$\underline{S}_3$	$\underline{S}_4$	$\underline{S}_5$
-8	-3.177	3.736	-5.117	4.003	-27.408
-7	-2.257	3.734	-12.166	28.431	220.437
-6	-1.917	3.733	-7.187	26.320	91.826
-5	-1.619	3.732	2.208	24.901	81.516*
-4	-1.089	3.732	-21.221	61.184*	-5.304*
-3	8.634	3.731	-12.892*	5.673*	-5.572
-2	-2.825	3.724*	-5.191*	9.285	-6.364
-1	-2.167*	3.701*	-4.477	7.083	-179.415
0	-1.856*	3.691	-1.372	7.124	-5.327
1	-1.547	3.690	-7.677	-6.694	-6.548
2	-.896	3.690	-5.323	3.634	-8.502
3	-12.228	3.690	-13.650	9.761	2.505
4	-2.613	3.689	-7.846	1.059	-13.610
5	-2.089	3.688	-7.124	-2.673	-49.582
6	-1.794	3.686	-7.723	39.953	4.982
7	-1.459	3.685	-6.926	7.759	-11.612

The operator $1 - 1.71B + B^2$ thus constructed is a nonstationary operator and, unless it is a repeated factor, Table XVIII suggests that the data transformed by this operator will be stationary. Thus let $w_t = (1-1.71B+B^2)X_t$. Table XIX gives the R and S-Arrays for w_t. Clearly from $S_2(m)$, $p = 2$, $q = 1$. The fact that $q = 1$ is also clear in $R_3(m)$. The model thus obtained is

$$\phi(B)w_t = (1-\theta B)Z_t, \tag{24}$$

where $\phi(B)$ is a stationary operator. Now from the IMSL subroutine FTMAXL one obtains from the w_t series (shown in Figure 7 on an <u>expanded</u> scale)

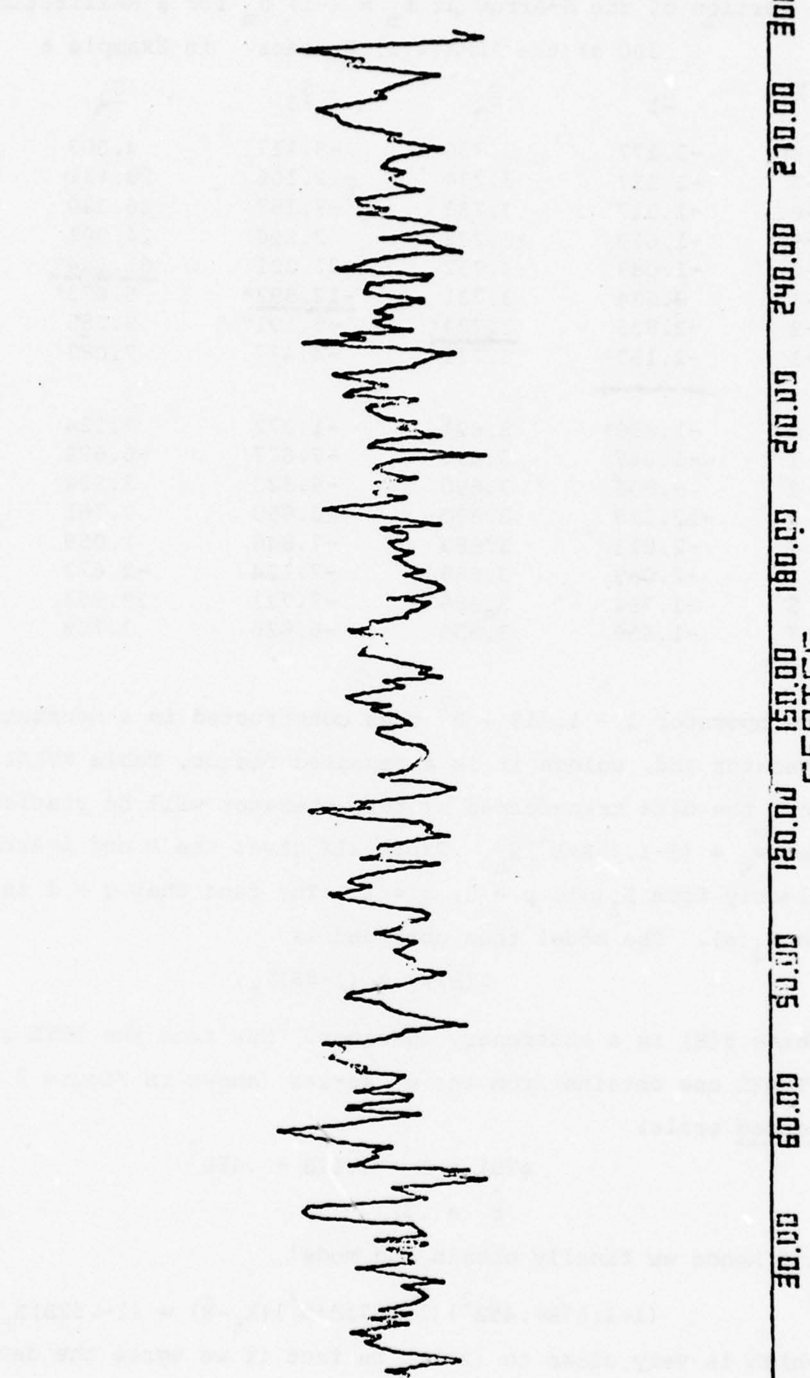
$$\hat{\phi}(B) = 1 - 1.17B + .45B^2$$
$$\hat{\theta} = .52$$

and hence we finally obtain the model

$$(1-1.17B+.45B^2)(1-1.71B+B^2)(X_t-\bar{X}) = (1-.52B)Z_t, \tag{25}$$

which is very close to (23). In fact if we agree the data should be modeled as nonstationary, this model is about as good as one could expect to get.

Figure 7



SERIES
30.00
60.00
90.00
120.00
150.00
180.00
210.00
240.00
270.00
300.00
12.00
6.80
1.60
-3.60
-8.80
-14.00

TABLE XIX

Portions of the R and S Arrays at $f_m = (-1)^m \hat{\rho}_m$ for the Series $w_t = (1-1.71B+B^2)X_t$, where X_t is the Series of Figure 5

	R-Array				S-Array		
k	R_1	R_2	R_3	k	S_1	S_2	S_3
-8	-.195	.001	-.026	-8	-1.952	-24.780	-1.064
-7	.186	.023	-.016	-7	-1.970	2.909	-7.655
-6	-.181	-.016	.005	-6	-1.717	5.878	41.254
-5	.129	.047	.036	-5	-1.503	6.501	-3.918
-4	-.065	-.683	.013	-4	-.419	6.152	48.943
-3	-.037	-.156	.129	-3	-7.971	5.627	19.511*
-2	.264	.112	.337*	-2	-3.272	12.102*	-1.686*
-1	-.601	-.398*	.029	-1	-2.661*	1.846*	-2.190
0	1.000*	.060	.005	0	-1.601*	2.732	5.600
1	-.601	-.054	.019	1	-1.440	2.442	-2.778
2	.264	.061	.002	2	-1.143	3.193	-21.343
3	-.037	.335	-.013	3	.722	3.216	1.094
4	-.065	-.025	-.017	4	-2.987	4.953	.390
5	.129	.027	-.020	5	-2.394	1.898	-3.788
6	-.181	.001	-.024	6	-2.030	-19.886	-8.893
7	.186	.019	.015	7	-2.049	5.495	-5.360
8	-.195	-.037	-.009	8	-1.849	3.294	-.412

Example 7.

Table XX gives the S-Array at $f_m = \hat{\rho}_m$ to 4 Columns from a realization of length 300 from an AR process. Since $S_3(m)$ is essentially constant the process is nonstationary with at least three roots on the unit circle. Since $S_3(m) \not\approx 0$, +1 is not a root. Thus we have at least one root of -1 and two complex roots on the unit circle. Hence if we first let $w_t = X_t + X_{t-1}$ and evaluate the S-Array for w_t, the complex operator for stationarity is easily found as in the previous example and the final factor is obtained from the resulting stationary series. The actual process for this realization is

$$(1-G_1B)(1-G_2B)(1-G_3B)(1-G_4B)X_t = Z_t, \tag{26}$$

where

$$G_1 = -.98,\ G_2 = .6,\ G_3 = .7 + .7i,\ G_4 = .7 - .7i$$

and hence the validity of our analysis is clear.

TABLE XX

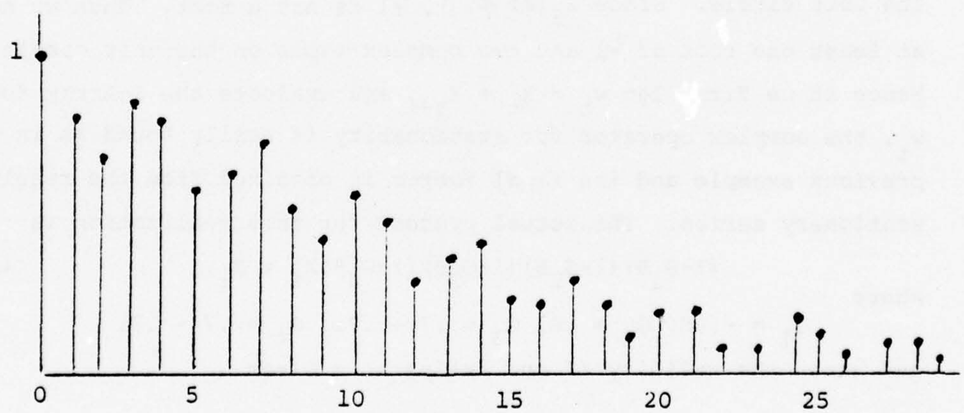
A Portion of the S-Array at $f_m = \hat{\rho}_m$ for a Realization of Length 300 from the AR(4) Process $\Pi_{i=1}^{4}(1-G_iB)X_t = Z_t$ where $G_1 = -.98$, $G_2 = .6$, $G_3 = .7+.7j$, $G_4 = .7-.7j$

k	$\underline{S}_1$	$\underline{S}_2$	$\underline{S}_3$	$\underline{S}_4$
-10	8.8240	.6895	-1.2095	.5710
-9	.3983	.5937	-1.1984	.2714
-8	-.3445	.4755	-1.2022	2.3810
-7	-1.0536	.7379	-1.1975	-3.9603
-6	21.1401	.4903	-1.1944	-.7487
-5	.3250	.5849	-1.2075	-3.5518
-4	-.2808	.6935	-1.1784	-1.5304*
-3	-1.0689	.4218	-1.2420*	.6201*
-2	14.2091	.7097*	-1.0425*	.8323
-1	.4601*	.5667*	-1.1233	.4306
0	-.3151*	.4468	-1.1084	.8622
1	-.9342	.7213	-1.1252	2.2520
2	-15.4989	.4718	-1.1223	-.3280
3	.3904	.5741	-1.1193	-1.0177
4	-.2453	.6628	-1.1249	-2.1680
5	-.9548	.4323	-1.1182	.6945
6	-19.6517	.6786	-1.1486	.8472

Example 8.

Figure 8 below shows a graph of the sample acf of a realization of length 300 from an ARMA process.

Figure 8



$\hat{R}(K)$

Table XXI shows the S-array at $f_m = (-1)^m \hat{\rho}(m)$ for the data.

TABLE XXI

m	S_1	S_2	S_3	S_4	S_5
-6	-1.901	.649	-3.211	12.855	-5.575
-5	-2.389	4.077	-3.010	7.945	60.297*
-4	-2.068	-.627	-3.116	8.529*	-4.163*
-3	-1.815	2.492	-3.795*	4.472*	8.970
-2	-2.181	-19.321*	-2.933*	4.752	-1.963
-1	-2.234*	1.654*	-2.481	3.812	-6.840
0	-1.809*	17.868	-2.888	4.954	-8.589
1	-1.846	.497	-2.919	5.765	-9.296
2	-2.227	3.489	-2.783	6.196	50.935
3	-1.936	-.787	-2.814	6.389	-5.375
4	-1.719	2.215	-2.951	4.856	-25.993
5	-2.109	-43.331	-2.794	4.687	-7.443

An inspection of the table immediately suggests from Column 1 that the data needs differenced. Moreover from Columns 3 and 4 one would assume (by Column 3) that the process also has two complex roots on the unit circle and is a fourth order process. However, before making decisions regarding Colums 3 and 4 one should first eliminate the simplest indicated nonstationarity. That is, the data should be differenced. Upon differencing the data one obtains Table XXII taking $f_m = \hat{\rho}_m$ which in this case yielded a more obvious pattern than $f_m = (-1)^m \hat{\rho}_m$.

TABLE XXII

S-ARRAY AT $f_m = \hat{\rho}_m$ FOR $(1-B)X_t$ IN EXAMPLE 8

m	S_1	S_2	S_3
-5	-1.573	2.535	1.787
-4	.331	2.522	2.088
-3	-2.275	2.727	1.085*
-2	-.753	2.578*	-.730*
-1	-6.103*	2.230*	-1.103
0	-1.195*	2.464	-1.006
1	3.050	2.543	-1.297
2	-1.783	2.427	-.927
3	-.248	2.433	.047
4	-2.744	2.465	.632

Inspection of Table XXII, then suggests the process has two complex roots on the unit circle. Proceeding now as in Example 6 gives at once the operator $1 + .51B + B^2$ as a nonstationary factor. Applying this operator to the differenced data then yields the R and S Arrays of Table XXIII.

TABLE XXIII

S AND R-ARRAYS FOR $(1+.51B+B^2)(1-B)X_t$ IN EXAMPLE 8

$S_1(m)$	$S_2(m)$	m	$R_1(m)$	$R_2(m)$
-2.494	.632	-5	-.102	.013
-2.718	-1.518	-4	.152	.010
-2.609	-11.483	-3	-.262	.016
-2.712	9.999*	-2	.421	.087
-2.385*	2.080*	-1	-.721	-.278*
-1.721*	1.311	0	1.000*	.057
-1.584	.581	1	.721	.009
-1.621	-.505	2	.421	.006
-1.581	2.891	3	-.262	.008
-1.669	5.530	4	.152	-.006
-1.770	-.332	5	-.102	.013

Finally by the D-Statistic or visually analyzing Table XXIII one can see the resulting data is ARMA (1,1) and the final model is

$$(1-.56B)(1+.51B+B^2)(1-B)X_t = (1+.4B)Z_t.$$

The coefficient .56 is obtained from Table XXIII from the ratio

$$\frac{S_1(1)}{S_1(-2)} = \frac{1.584}{2.712} = .56$$

and as mentioned before is the conditional Yule-Walker estimate (i.e., given p <u>and</u> q). The estimate .4 on the right side of the equation was obtained by standard methods. Since the data of this example was generated from the model

$$(1-.6B)(1+.5B+.98B^2)(1-.98B)X_t = (1+.4B)Z_t$$

the above analysis appears quite good.

Example 9.

Before returning to real data let us consider one more set of simulated data. Again we consider an AR(4) process. Figure 9 shows 300 generated points from the process

$$(1-G_1B)(1-G_2B)(1-G_3B)(1-G_4B)X_t = Z_t, \qquad (27)$$

where

$$G_1 = -.99,\ G_2 = .98,\ G_3 = .5+.4i,\ G_5 = .5-.4i.$$

Table XXIV shows the S-Array at $f_m = \hat{\rho}_m$ and $f_m = (-1)^m\hat{\rho}_m$.

Figure 9

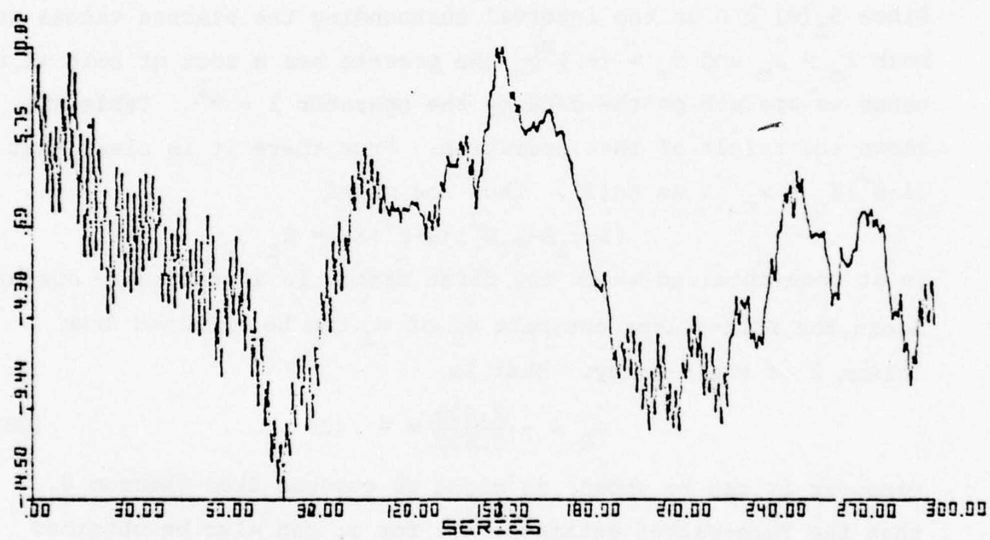


TABLE XXIV

S-Arrays

$f_m = \hat{\rho}_m$				$f_m = (-1)^m \hat{\rho}_m$		
$S_1(m)$	$S_2(m)$	$S_3(m)$	m	$S_1(m)$	$S_2(m)$	$S_3(m)$
.266	-.073	.015	-7	-2.266	-.045	.033
-.149	-.077	.110	-6	-1.850	.004	.039
.259	-.082	-.726	-5	-2.259	-.037	.419
-.136	-.080	-.165	-4	-1.863	-.052	.055
.249	-.064	-.074*	-3	-2.249	.094	.188*
-.131	-.033*	-.045*	-2	-1.868	-.346*	-.114*
.214*	.028*	-.054	-1	-2.214*	.291*	-.018
-.176*	.065	-.069	0	-1.823*	-.095	-.040
.151	.075	-.286	1	-2.151	.049	.101
-.199	.077	.018	2	-1.800	.035	-.037
.157	.074	-.057	3	-2.157	-.004	-.042
-.206	.069	-.393	4	-1.793	.042	.166
.175	.068	.074	5	-2.175	.035	-.074
-.210	.074	.016	6	-1.789	-.001	-.075
.183	.079	-.013	7	-2.183	.076	-.046

Since $S_2(m) \approx 0$ in the interval surrounding the starred values at both $f_m = \hat{\rho}_m$ and $f_m = (-1)^m \hat{\rho}_m$ the process has a root of both ± 1 and hence we operate on the data by the operator $1 - B^2$. Table XXV shows the result of that operation. From there it is clear that $(1-B^2)X_t = w_t$ is an AR(2). Thus the model

$$(1-\psi_1 B-\psi_2 B^2)(1-B^2)X_t = Z_t$$

is at once obtained where the first factor is a stationary operator. Again the Yule-Walker estimate $\hat{\psi}_2$ of ψ_2 can be obtained from Column 2 of the S-array. That is

$$\hat{\psi}_2 = -\frac{2.439}{5.690} = -.429 . \tag{28}$$

Moreover it can be shown, as might be guessed from Theorem 4, that the Yule-Walker estimate, $\hat{\psi}_1$, for ψ_1 can also be obtained from $S_2(m)$. That is

$$1 + \hat{\psi}_1 - \hat{\psi}_2 = 2.439$$

or

$$\hat{\psi}_1 = 1.01 \tag{29}$$

TABLE XXV

S-Array for $w_t = (1-B^2)X_t$ at $f_m = (-1)^m\hat{\rho}_m$ in Example 9

	$S_1(m)$	$S_2(m)$	$S_3(m)$
-7	-12.445	3.802	-221.284
-6	-2.438	3.875	-1.360
-5	-1.675	7.011	-10.735
-4	-.655	6.539	9.207
-3	-11.885	7.166	-27.116*
-2	-3.471	5.690*	-2.680*
-1	-2.413*	2.439*	-1.308
0	-1.707*	2.117	-7.002
1	-1.404	2.327	.719
2	-1.091	2.737	-3.084
3	1.902	2.986	-12.297
4	-2.481	3.030	2.067
5	-1.695	3.202	.729
6	-1.087	3.048	-8.949
7	7.595	2.918	-3.245

Thus the estimated model is

$$(1-1.01B+.429B^2)(1-B^2)X_t = Z_t$$

whereas the true model is

$$(1-B+.41B^2)(1-.97B^2)X_t = Z_t .$$

Of course the estimates (28) and (29) can be obtained by more sophisticated methods but in this example its not necessary.

Example 10. (Series A)

Consider now the 197 data points given in Box and Jenkins and referred to as Series A. The data consists of concentration readings taken every two hours from a chemical process. A graph of the data (connected by straight lines) is given in Figure 10. The data are shown in tabular form in [5] where it was modeled as either the stationary ARMA (1, 1) model

$$X_t - .92X_{t-1} = 1.45 + a_t - .58a_{t-1} , \tag{30}$$

or the nonstationary ARIMA (0, 1, 1) model

$$(1-B)X_t = a_t - .70a_{t-1} . \tag{31}$$

In Tables XXVI and XXVII the S-array at $f_m = \hat{\rho}_m$ and $f_m = (-1)^m\hat{\rho}_m$ are given. At $f_m = (-1)^m\hat{\rho}_m$ there is clear support in Table XXVII for the models (30) and (31). However we should remember that

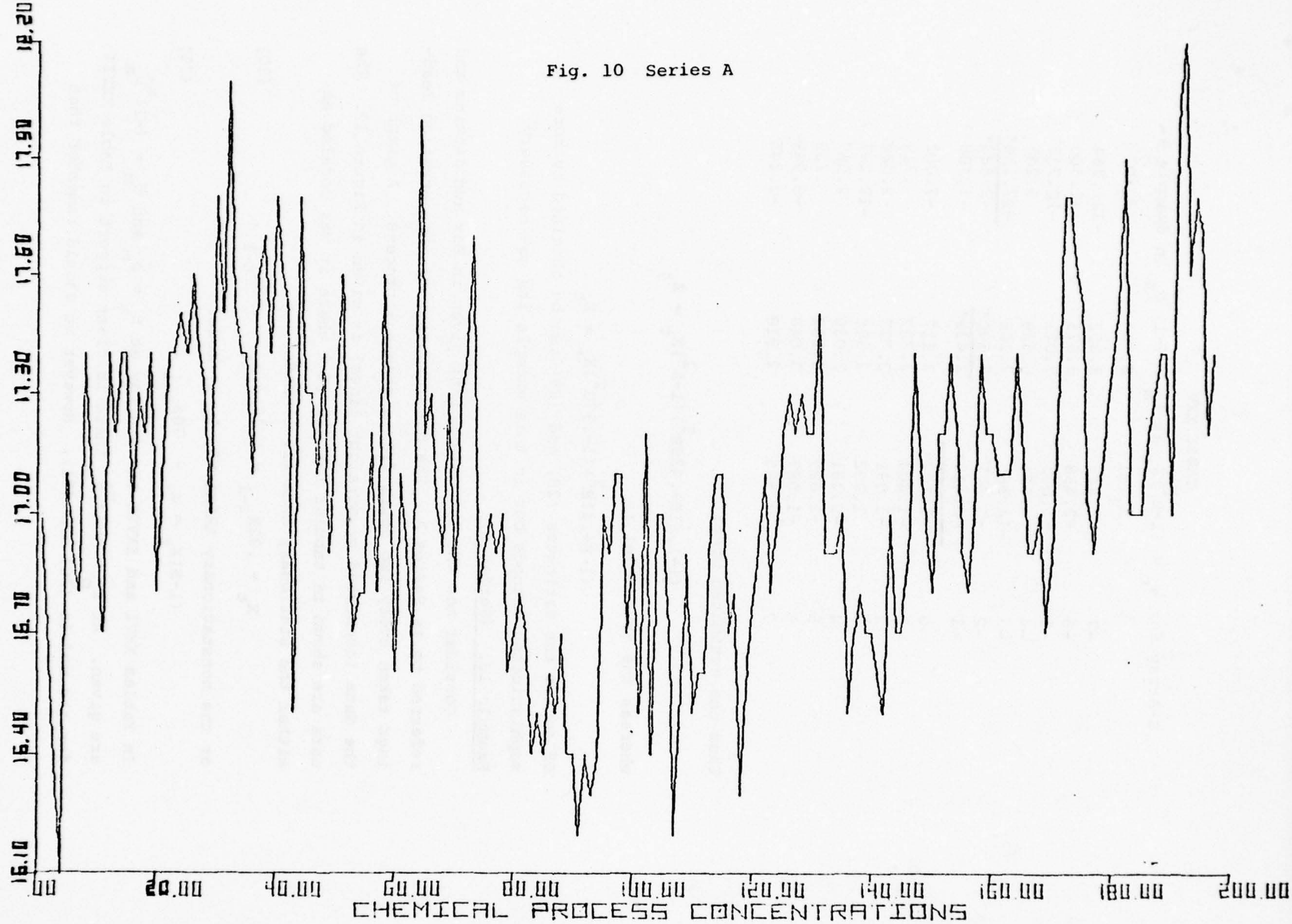


Fig. 10 Series A

TABLE XXVI

S-Array for Series A at $f_m = \hat{\rho}_m$

m	S_1	S_2	S_3	S_4	S_5	S_6	S_7	S_8
-10	.193	-.299	-1.070	-.054	.167	-.377	1.330	-1.633
-9	.057	-.359	-.009	-.075	.346	14.114	1.303	-15.787
-8	.218	.082	.062	-.235	.416	-2.904	1.315	6.265*
-7	-.107	-.301	.196	-1.415	4.619	2.295	1.232*	.198*
-6	-.065	-.102	.288	-.136	4.092	-1.846*	-.192*	.257
-5	.088	.284	.028	.909	3.959*	.228*	-.237	1.216
-4	.118	-.352	.134	-4.025*	-.260*	-.095	-.236	-1.128
-3	.244	-1.984	4.386*	.278*	.261	.652	-.245	.128
-2	.151	-1.276*	-.299*	-.004	.266	-.290	-.272	.642
-1	.753*	.321*	.024	-.319	.385	2.847	-.290	-.304
0	-.429*	.212	.004	-.259	.298	.204	-.345	.603
1	-.131	.344	.253	.923	-.234	.826	-.554	-.748
2	-.196	.057	.206	-.095	-.218	.298	-.588	.245
3	-.106	-.504	-.152	.128	-.161	.105	-.570	
4	-.081	-.098	.017	.153	.557	.367	-.588	
5	.069	-.586	-.149	.214	-.263	1.633		

TABLE XXVII

S-Array for Series A at $f_m = (-1)^m \hat{\rho}_m$

m	$S_1(m)$	$S_2(m)$	$S_3(m)$	$S_4(m)$	$S_5(m)$	$S_6(m)$	$S_7(m)$	$S_8(m)$
-10	-2.193	-.698	6.166	1.847	-3.862	-13.251	-4.558	4.429
-9	-2.057	-2.839	-2.214	6.957	12.534	-538.187	-5.558	-55.909
-8	-2.218	1.637	-3.631	-2.392	11.817	-55.478	-5.783	41.030*
-7	-1.892	8.309	.119	-1.674	6.845	-63.183	-8.071*	1.301*
-6	-1.934	2.508	2.360	-12.148	53.718	-8.822*	-1.261*	.913
-5	-2.088	11.663	13.038	-498.648	-18.922*	1.090*	-1.045	-3.299
-4	-2.118	.800	13.128	-16.860*	-1.244*	2.640	-1.008	6.694
-3	-2.244	-14.706	-18.371*	1.168*	-3.430	12.466	-.840	2.388
-2	-2.151	-4.604*	-1.254*	2.354	-.305	11.069	-1.893	-1.175
-1	-2.753*	1.174*	-2.440	-28.418	-10.928	100.068	-1.261	2.929
0	-1.570*	1.573	-2.292	-.306	-10.826	4.311	-.110	2.637
1	-1.868	-.293	-2.079	9.380	-5.409	-10.062	-2.653	9.108
2	-1.803	2.344	-.125	8.759	-3.972	-.461	-2.706	4.537
3	-1.893	12.357	-8.910	-4.335	-2.027	1.959	-3.043	-24.962
4	-1.918	2.755	-4.187	1.767	-10.298	6.800	-2.707	2.661

For this Data D(1, 1) = 6.92 while D(7, 1) = 440.18

$S_1(m)$ can take on the characteristics shown there strictly because of the near or actual nonstationarity. Therefore a close inspection of the S-arrays is in order. Table XXVI, although supportive of (30) and (31) is not nearly so distinct as Table XXVII in this regard. Note that there is a very distinctive behavior in Table XXVI in Columns 7 and 8. That is, from those columns the process appears to be either an ARMA (7, 1) or ARMA (7, 0), and another inspection of Table XXVII supports this. Now as stressed before, the decision to include the moving average term when modeling the process as a nonstationary model should be made after the nonstationarity is removed. Moreover the only thing that Table XXVI clearly indicates is that the data should be differenced. Table XXVIII shows the S-array of the differenced data at $f_m = \hat{\rho}_m$ and the process is clearly ARMA (6, 0). An R-array Table of the data is equally distinct. A preliminary estimate of the coefficients in the models (corresponding to (30) and (31))

$$(1-\phi_1B-\phi_2B^2-\phi_3B^3-\phi_4B^4-\phi_5B^5-\phi_6B^6-\phi_7B^7)(X_t-\bar{X}) = a_t-\theta_1a_{t-1} \qquad (32)$$

and

$$(1-\psi_1B-\psi_2B^2-\psi_3B^3-\psi_4B^4-\psi_5B-\psi_6B^6)(1-B)X_t = a_t \qquad (33)$$

gives the Yule-Walker estimates

$$\hat{\phi}_1 = .37 \quad \hat{\phi}_2 = .2 \quad \hat{\phi}_3 = .02 \quad \hat{\phi}_4 = .01 \quad \hat{\phi}_5 = -.01 \quad \hat{\phi}_6 = .06 \quad \hat{\phi}_7 = .16$$

$$\hat{\psi}_1 = -.61 \quad \hat{\psi}_2 = -.40 \quad \hat{\psi}_3 = -.36 \quad \hat{\psi}_4 = -.32 \quad \hat{\psi}_5 = -.31 \quad \hat{\psi}_6 = -.21 \ .$$

These sets of estimates are compatible and suggest the parsimonious models,

$$(1-\phi_1B-\phi_2B^2-\phi_7B^7)(X_t-\bar{X}) = a_t- \Theta\ a_{t-1} \qquad (34)$$

or

$$(1-\psi_1B-\psi_2B^2-\psi_2B^3-\psi_2B^4-\psi_2B^5-\psi_2B^6)(1-B)X_t = a_t \ . \qquad (35)$$

Note that like (32) and (33), (34) and (35) are compatible, i.e., $\psi_2 = \psi_3 = \psi_4 = \psi_5 = \psi_6$ implies $\phi_3 = \phi_4 = \phi_5 = \phi_6 = 0$. In addition it should be noted that the nonstationary ARIMA (6, 1, 0) in (35) is parsimonious in that there are only two parameters to estimate. Any one of the models (32), (33), (34) or (35) yield a slightly smaller

TABLE XXVIII

S-Array for the Differences of Series A at $f_m = \hat{\rho}_m$

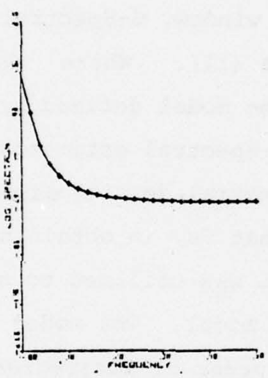
m	S_1	S_2	S_3	S_4	S_5	S_6	S_7	S_8
-10	.770	2.706	2.946	3.084	26.596	14.616	-23.326	19.620
-9	-2.830	-35.245	-2.396	2.062	2.268	15.042	12.052	45.835
-8	-2.148	1.908	2.542	-3.213	-50.243	14.697	-14.341	72.290*
-7	-1.139	1.144	-.408	-2.943	-5.145	15.028	-1522.851*	3.368*
-6	2.378	.148	-2.854	-.743	-6.228	15.017*	-3.218*	2.097
-5	-.841	.418	3.188	4.906	-13.740*	3.212*	67.601	5.357
-4	5.059	-3.650	-37.704	15.955*	-2.646*	3.187	-1.441	6.427
-3	-1.281	-2.482	-11.786*	2.218*	-.273	3.238	-8.649	10.523
-2	-23.209	9.127*	-1.947*	-.426	-7.866	3.199	-2.126	6.454
-1	-3.421*	1.671*	-.579	.054	-2.104	3.071	-22.711	9.160
0	-1.412*	.439	-2.792	2.136	7.486	3.044	1.651	6.459
1	-1.045	.617	.177	2.228	-3.547	4.253	-6.703	-7.393
2	-4.547	-.404	-4.207	3.119	-8.947	8.724	-1.370	-10.688
3	-.834	-.151	-2.508	3.822	-3.618	7.219	7.235	
4	5.297	2.655	-6.964	1.259	-1.993	7.604	18.346	
5	-.703	3.637	-.861	-2.159	-3.213	7.025		

estimate of the white noise variance than (30) or (31). This is not particularly significant in view of the models even though (34) and (35) are reasonably parsimonious.

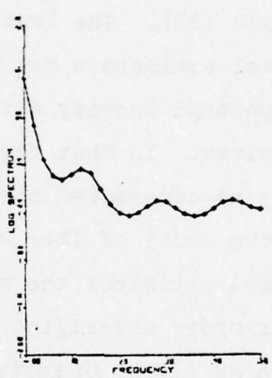
There are however a number of observations which can be made concerning Series A which suggest that the models (34) or (35) are more appropriate models for the data. In particular a spectral analysis of the data by four different methods, standard window methods, the AR Spectral Estimator, the Maximum Entropy Method, and the G-Spectral Estimator all show a maximum peak in the spectrum at the frequency w = 0 as well as a local peak near w = .12, while all but the windows show another small peak near the w = .26. This of course suggest the autocorrelation is not that of a process such as (30) and (31). The logs of the Parzen window, G-Spectral and AR-Spectral estimators are shown in Figure (11). Where the log spectral density associated with the model defined by (30) is also given. In that figure the log AR-Spectral estimator can essentially be considered the log of the spectral density associated with the model of (34) with $\theta = 0$. That is, in obtaining the AR-Spectral estimator the method of Akaike was utilized to obtain the proper order underlying autoregressive model. The model selected was an AR (7, 0) in agreement with the order of autoregressive in (34), and the estimated coefficients were essentially the $\hat{\phi}$'s already given. Thus it is clear that the estimated spectral density obtained in a variety of ways suggests the model (34) and (35) as more appropriate than (30) or (31). In addition that analysis suggests the autocorrelation very nearly has a damped periodic component whose period is in the neighborhood of 7 or 8 lags. It is quite possible that such observations as this may be extremely important to the experimenter and they are completely lost in both of the models (30) and (31). Of course, there is no absolute answer here as to whether the models of (34) and (35) are better than those of (30) and (31) since the true model is not known. To shed some additional light on the methods of this paper, the reasons why a model such as (34)

or (35) might be identified as (30) or (31) or vice-versa were investigated by a short Monte Carlo study. The result of simulating 5 realizations of (30) of length 197 was that the ARMA (1,1) model was always selected via the R and S array technique. On the other hand when realizations of (34) of the same length were simulated the ARMA (7,0) or ARMA (7,1) was always chosen. Typical S arrays for the simulated ARMA (1,1) and ARMA (7,1) are shown in Tables XXIX and XXX. From these tables the ARMA (1,1) model of (30) is clearly distinguishable from the model of (34), and hence these authors feel that (34) or (35) is more appropriate for series A and nearly as parsimonious as the (1,1) model.

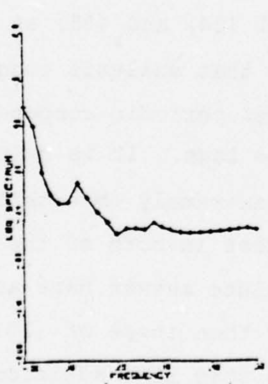
Fig. 11 Spectral Estimators for Series A



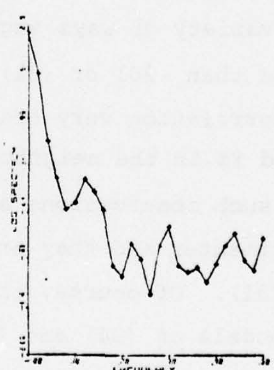
Spectral Density from (30)



AR-Spectral Estimator



G-Spectral Estimator



Parzen Window Spectral Estimator

TABLE XXIX

S-Array for a Realization of Length 197 from the Process of Equation (30) at $f_m = (-1)^m \hat{\rho}_m$

m	$S_1(m)$	$S_2(m)$	$S_3(m)$	$S_4(m)$	$S_5(m)$	$S_6(m)$	$S_7(m)$	$S_8(m)$
-10	-1.949	.977	3.198	-2.231	-.442	-6.313	-78.800	-17.839
-9	-2.269	-4.609	-6.110	-.567	3.250	-6.236	20.915	26.576
-8	-2.102	-2.083	.063	-.661	-3.280	-5.867	15.400	38.816*
-7	-2.550	.387	.501	-2.006	3.974	-4.000	19.284*	.860*
-6	-1.913	-.283	1.940	-329.115	8.109	-11.197*	-.841*	.746
-5	-2.331	-.620	2.028	-7.585	-54.046*	.879*	-1.287	2.261
-4	-1.902	.569	5.199	-6.415*	-.955*	.354	-1.178	-3.161
-3	-2.416	-5.947	-24.542*	.938*	-5.114	.953	-.790	-.261
-2	-2.064	-3.741*	-1.099*	1.069	.364	.847	18.481	-.798
-1	-3.162*	1.052*	-4.517	.331	-.307	.851	-1.079	-.728
0	-1.463*	1.267	-.336	-34.965	-1.293	1.091	-75.250	1.340
1	-1.940	-.649	-.333	.495	-.413	1.089	-.909	5.043
2	-1.706	.430	.958	1.030	-1.183	-3.187	-1.246	.178
3	-2.108	.227	-.120	.942	-.495	-1.025	-.611	-.074
4	-1.752	-.418	-.916	.236	-2.447	.060	-.635	.731
5	-2.095	.858	-3.522	.444	.589	.589	-.275	0.000

TABLE XXX

S-Array at $f_m = (-1)^m \hat{\rho}_m$ for a Realization of Length 197 from the Process of (34) with $\phi_1 = .42$, $\phi_2 = .17$, $\phi_7 = .26$ and $\theta = -.26$

m	$S_1(m)$	$S_2(m)$	$S_3(m)$	$S_4(m)$	$S_5(m)$	$S_6(m)$	$S_7(m)$	$S_8(m)$
-10	-2.132	13.024	-113.038	.218	-3.511	-121.817	-7.006	2.308
-9	-2.118	9.309	-2.146	3.544	-.470	-103.037	-6.526	42.023
-8	-2.051	1.046	-3.947	8.522	109.043	165.144	-6.226	44.122
-7	-1.832	3.323	-2.149	4.709	51.198	67.443	-8.042*	2.000
-6	-1.933	5.210	-7.090	-153.438	-123.032	-11.217*	-1.913*	2.022
-5	-2.014	4.114	.477	-156.911	-14.759*	1.545*	-1.635	-1.014
-4	-2.147	3.430	137.619	-19.250*	-1.792*	4.582	-1.723	6.879
-3	-2.260	4.685	-44.648*	1.598*	-3.579	-16.082	-1.855	.194
-2	-2.309	-45.912*	-1.743*	4.130	3.910	-41.875	-1.486	.892
-1	-2.348*	1.677*	-5.306	8.073	21.067	29.139	-1.215	-2.518
0	-1.741*	3.536	-.507	7.305	-2.440	33.427	-1.154	3.688
1	-1.763	5.487	-8.046	-21.977	-29.255	.194	-1.191	-.800
2	-1.793	4.153	3.693	-41.751	-2.305	1.236	1.030	11.909
3	-1.871	3.377	16.435	-.277	-2.096	1.824	.756	1.724
4	-1.985	5.261	-4.742	2.146	1.254	2.377	-1.802	-8.862
5	-2.071	-2.598	-2.452	-1.344	-72.371	3.358	-1.234	0.000

VII. SMALL SAMPLE SIZES

In the previous examples the number of observations was always 100 or more. This was done so that patterns would be very distinct to the reader attempting to understand the method. However, once the method is understood, processes which are reasonably modeled by ARMA models can be identified on much smaller sample sizes. Although it is realized that the number of samples required to accurately identify p and q would vary dramatically with the coefficients, in order to obtain some idea of how small the sample size might be and still yield clear identification of p and q, the authors generated 3 realization of each of the processes in the previous examples at sample sizes 25, 50, 75, 100, 150 and 200. In all but one case, Example 6 which is discussed below, p was clearly correctly identifiable in sample sizes as small as 25 in at least two of the three realizations. In most cases q was either correctly identified or identified as q + 1 or q + 2 in samples of size 25. The patterns in the R & S array were not markedly better for samples of size 50 than for samples of size 25. However, by 75 or 100 samples the patterns were usually very good and with a few exceptions all the realizations were correctly identified, even by the D-Statistic. Although the D-Statistic is not recommended alone for small samples, these authors found it very useful even then. That is, in each realization D(n,m) was calculated with Max n = 4 and Max m = 2. By investigating the R and S arrays at all large values of D(n,m) the most predominate patterns were considered.

In order to substantiate the claims made in the above, three examples are now given.

Example 11.

Let us consider once more the process

$$X_t - 1.32X_{t-1} + .68X_{t-2} = Z_t - .8Z_{t-1} \qquad (36)$$

of Example 1.

Table XXXI shows the S and R-arrays using $f_m = (-1)^m\hat{\rho}_m$ for a sample of size 25 generated from (36) with Z_t i.i.d. N(0,1) while Table XXXII show the calculated values of D(n,m). Clearly from Table XXXI the process is ARMA (2,1). Moreover inspection of D(n,m) in Table XXXII also identifies the process as ARMA (2,1). As was mentioned, 3 realizations were generated at each of the previously listed sample sizes. For samples of size 25 in this example, one realization yielded a much more distinct ARMA (2,1) pattern than those of Table XXXI and one yielded a much less distinct pattern. When the sample size was as large as 100 all realizations were easily identifiable.

TABLE XXXI

S and R Array for Example 11 with $f_m = (-1)^m\hat{\rho}_m$ for a Sample of Size 25

$S_1(m)$	$S_2(m)$	$S_3(m)$	m	$R_1(m)$	$R_2(m)$	$R_3(m)$
-4.049	4.667	2.151	-6	-.082	-.103	-.032
-2.383	3.168	-7.168	-5	.251	.099	-.053
-1.704	5.158	84.758	-4	-.347	-.085	.021
-1.112	5.505	5.958*	-3	.245	.309	.341
11.343	9.595*	-1.235*	-2	-.028	.459	.552*
-3.936*	1.558*	2.896	-1	-.341	-.659*	.115
-1.340*	3.224	-13.301	0	1.000*	.107	.012
-.918	3.084	-1.097	1	-.341	-.269	-.039
-9.867	3.728	4.532	2	-.028	-.185	-.027
-2.419	2.714	4.384	3	.245	.100	-.073

TABLE XXXII

D(n,m) From a Sample of Size 25 in Example 11 at $f_m = (-1)^m\hat{\rho}_m$

m \ n	0	1	2	3	4
0	8.391*	.011	.143	.006	.158
1	30.750*	.000	23.970	.001	1.852
2	.014*	2.791	.292	.042	.000
3	.712*	.056	.103	.069	.001

The symbol * in D(n,m) tables is to remind the reader that n=0 is not included in the D-Statistic since these values are not scaled the same as other values of n.

It should be noted that the pattern in the S-array is much more clear if one focuses on the bottom part of $S_2(m)$. From Monte Carlo studies made by the authors, this is to be expected. That is, in some limited investigations the authors found $\hat{S}_p(-p+1)$, $\hat{S}_p(-p+2)$ ..., $\hat{S}_p(-p+n)$ to have lower mean square errors than their counterparts $\hat{S}_p(-p)$, $\hat{S}_p(-p-1)$ ..., $\hat{S}_p(-p+n-1)$, and hence for very small samples it is probably wise to concentrate on the region very close and below the starred values. In this example the D-Statistic clearly gives the proper estimate for p and q. Again, however, for samples this small it is probably better to use a table such as Table XXXII as simply a guide for visual inspection. Finally we should mention that neither this example nor the ones that follow are meant to suggest that 25 samples are usually enough data to determine p and q in an ARMA (p,q) process. They do, however, suggest that in many cases p and q can be accurately determined even for complex processes with a surprisingly small amount of data.

Example 12.

Consider now the process of Example 6, i.e.

$$(1-1.2B+.45B^2)(1-1.7B+.99B^2)X_t = (1-.5B)Z_t \tag{37}$$

The S and R arrays are given in Table XXXIII for a sample of size 25. The process is indicated there as an ARMA (2,1). Moreover it is clearly nearly nonstationary. Proceeding as before one obtains the operator $1-1.69B+B^2$ as the proper transformation to stationarity. Upon applying this operator to the data to obtain $V_t = X_t - 1.69X_{t-1} + X_{t-2}$ only the moving average term remained identifiable, i.e. the process appeared as an MA (1). This is not surprising since, as a review of Example 6 shows, the nonstationary factor is by far the more dominant feature of this model. Although one of the three realizations of length 50 clearly yielded the proper nonstationary ARMA (4,1) model, a sample of size 200 was required before this total pattern, i.e. both factors in (37), appeared with any consistency.

TABLE XXXIII

S and R Arrays for a Sample of Size 25

for Example 12 at $f_m = (-1)^m \hat{\rho}_m$

$S_1(m)$	$S_2(m)$	$S_3(m)$	m	$R_1(m)$	$R_2(m)$	$R_3(m)$
-1.999	3.845	-13.475	-6	-.772	-.125	.000
-1.675	3.841	5.553	-5	.771	.162	.000
-1.154	3.838	39.622	-4	-.520	-.377	.001
4.294	3.833	7.697*	-3	.080	-.714	.009
-2.960	4.004*	-2.341*	-2	.423	.211	.027*
-2.207*	3.365*	-3.233	-1	-.828	-.171*	.008
-1.828*	3.559	-2.097	0	1.000*	.144	.001
-1.510	3.546	-4.960	1	-.828	-.196	.000
-.811	3.547	-35.552	2	.423	.660	.000
-7.513	3.546	-7.302	3	.080	.348	.001

Example 13.

In this example we consider a sample of size 25 from the process of Example 9, i.e.

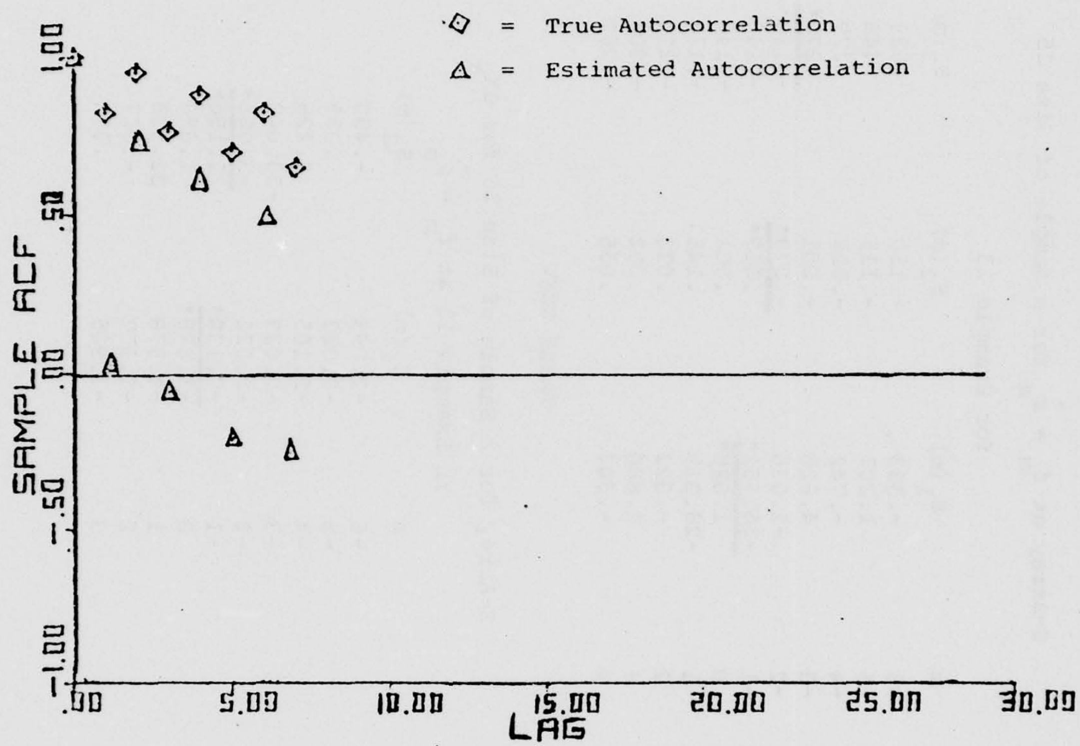
$$(1-\psi_1 B)(1-\psi_2 B)(1-\psi_3 B)(1-\psi_4 B) = Z_t, \qquad (38)$$

where

$$\psi_1 = -.99,\ \psi_2 = .98,\ \psi_3 = .5+4i,\ \psi_4 = .5-.4i.$$

The S-array at $f_m = (-1)^m \hat{\rho}_m$ is shown in Table XXXIV. The small values of $S_2(m)$ are characteristic of the factor $1-B^2$. Moreover when $f_m = (-1)^m \hat{\rho}_m$ the small tabular values in Column 3 strongly suggests -1 is a root of the characteristic equation. Performing the operator $1-B^2$ in the sequence $(1+B)(1-B)$ yields the S-array at $f_m = \hat{\rho}_m$ in Table XXXV for $\Delta X_t = X_t - X_{t-1}$ and the S-array for $W_t = X_t - X_{t-2}$ at $f_m = (-1)^m \hat{\rho}_m$ in Table XXXVI. Note that Table XXXV gives strong confirmation of the diagnosis that $1-B^2$ is a factor of $\phi(B)$, i.e. for ΔX_t, $S_1(m) \approx -2$ at $f_m = \hat{\rho}_m$. But Table XXXVI then identifies the resulting data as AR (2,0). This can also be seen from the table of values of D(n,m) in Table XXXVII. A plot of the true autocorrelation and the estimated autocorrelation is shown in Figure 12 out to seven lags which is the maximum number required to calculate the $S_2(m)$ column of Table XXXIV. It is interesting to note that although the estimates are poor their

Figure 12



Autocorrelations in Example 13

general pattern remains that of the true autocorrelation and hence explains why p and q are clearly distinguishable in this example.

TABLE XXXIV

S-Array at $f_m = \hat{\rho}_m$ for a Sample of Size 25 for Example 13

m	$S_1(m)$	$S_2(m)$	$S_3(m)$
-6	-.393	-.150	.639
-5	1.207	-.115	.448
-4	-.742	-.222	.279
-3	4.628	-.001	.279*
-2	-1.036	-.278*	-.122*
-1	-36.451*	.219*	-.146
0	1.028*	.001	-.145
1	-28.919	.145	-.112
2	-.822	.073	-.210
3	2.880	.092	-.308
4	-.547	.055	-.262

TABLE XXXV

S-Array for a Sample of Size 25 for ΔX_t in Example 13 at $f_m = \hat{\rho}_m$

m	$S_1(m)$	$S_2(m)$
-6	-2.093	-.490
-5	-2.023	.029
-4	-2.105	2.226
-3	-2.029	-33.049
-2	-2.022	-3.001*
-1	-2.139*	1.156*
0	-1.878*	1.858
1	-1.978	21.525
2	-1.972	-.027
3	-1.905	.373

TABLE XXXVI

S-Array for a Sample of Size 25 For

$W_t = X_t - X_{t-2}$ in Example 13 at $f_m = (-1)^m \hat{\rho}_m$

m	$S_1(m)$	$S_2(m)$	$S_3(m)$	$S_4(m)$
-6	-2.309	-4.028	-9.327	13.404
-5	-2.799	6.888	1.022	13.036
-4	-1.993	4.197	.386	13.260*
-3	-.572	4.519	9.343*	4.267*
-2	-6.613	5.707*	-1.792*	4.968
-1	-2.675*	2.218*	-.183	4.979
0	-1.597*	2.887	-.429	.469
1	-1.178	2.324	-3.981	.628
2	1.338	2.189	-.203	10.378
3	-2.007	1.041	-2.462	-10.649
4	-1.556	-.287	-3.029	-8.828

TABLE XXXVII

Values of D(n,m) in Example 13 For

$W_t = X_t - X_{t-2}$ at $f_m = (-1)^m \hat{\rho}_m$

m \ n	0	1	2	3	4
0	2.201*	.017	1.970	.002	.001
1	2.180*	.003	.008	.356	.000
2	.099*	.779	.002	.001	2.027

As is clear from the above tables an ARMA (4,2) is also a possible choice. However the ARMA (2,0) pattern does look better in Table XXXVI. Moreover with a sample this small one would undoubtly choose the ARMA (2,0).

VIII. SEASONAL DATA

Thus far we have not considered any examples where there was a strong seasonal component of rather long period such as in a model of the type

$$(1-B^{12})X_t = (1-\Theta B^{12})Z_t \qquad (38)$$

At first glance one might not expect the methods of this paper to be of particular use in this problem, due to the large values of p and q, unless a large data set is available. The extent to which this is true is not presently known to the authors. However, a number of modifications can be made to the methods of this paper which might allow a general approach to the seasonal problem. In case of (38) such modifications are easily found so that handling that particular case is of no difficulty. However the authors have yet to establish a general approach to the problem at this time. Some preliminary results have been established and can be found in Kelley (1977) where an interesting example of the so called "Colored Fox Data" (see Anderson, 1977) is given. Hopefully a comprehensive approach to the problem will appear in a later paper.

IX. THE MOVING AVERAGE PROCESS

Thus far only cases where $p > 0$ have been considered. This is not to suggest that the method is not useful if the data happens to come from an MA (q) process. In fact these authors have found that to be one of simpler cases to distinguish. However, one should be reminded that in the case of a MA (q) the S-array will have no fixed pattern and hence the method will direct one back to the R-array which in turn leads one to estimate q from $R_1(m)$, i.e. from the values of the autocorrelation. Thus at this point the method coincides with that of Box and Jenkins. The D(n,m) values at n = 0 are useful in aiding one in not overlooking the MA (q) possibility when inspecting the R and S arrays. However, only comparing them to other D(n,m) values would lead to the choice of a moving average too often.

X. NON-NORMAL NOISE

In all previous examples Z_t was i.i.d. N(0,1). Of course since the statistic of interest was $\hat{\rho}_m$ there was no loss in generality in using only unit variance. However, the question of how much the method would be effected if the noise were not normal is of interest. Thus the authors considered all of the previous examples again for samples of the same size as initially considered but with first exponential noise and then Cauchy noise. For these examples and sample sizes no detectable difference was found in the effectiveness of the method. For small samples this may not be the case but that was not investigated.

XI. CONCLUDING REMARKS

In this paper the authors have presented what they feel is a satisfactory approach to the problem of estimating p and q in ARMA (p, q) modeling problems. Several additional topics concerning this problem have been investigated but not considered here, the most promising being seasonal models and the transfer function model. Work in these areas will appear in later papers.

Some closing remarks should be made concerning the D-Statistic. The current definition may be a bit ad hoc and some modifications of the statistic may be necessary. At the present this was as good as the authors could do in the way of finding a single measure of the presence of the desired pattern. Undoubtly some improvement can be made there. In addition it seems that something similar to the D-Statistic could be done in the nonstationary case so that the data could be transformed to stationarity "rather automatically" if necessary.

Finally it should be mentioned that programs for calculating R, S and D(n,m) arrays are included in the Appendix so that application of the methods of this paper should not be difficult.

APPENDIX 1

<u>Proof of Theorem 3</u>: Since $\{X_t\}$ is stationary ARMA (p, q) $\rho_k \in L(p,\Delta)$ for $p \geq q + 1$. Then there exist constants $\phi_1, \ldots, \phi_p$ such that

$$\rho_k = \sum_{i=1}^{p} \phi_i \rho_{k-i} \quad \text{for } k \geq q + 1. \tag{A1}$$

Moreover

$$S_p(\rho_{-k}) = \frac{\begin{vmatrix} 1 & 1 & \cdots & 1 \\ \rho_{-k} & \rho_{-k+1} & \cdots & \rho_{-k+p} \\ \rho_{-k+1} & \rho_{-k+2} & \cdots & \rho_{-k+p+1} \\ \vdots & \vdots & & \vdots \\ \rho_{-k+p-1} & \rho_{-k+p} & \cdots & \rho_{-k+2p-1} \end{vmatrix}}{\begin{vmatrix} \rho_{-k} & \rho_{-k+1} & \cdots & \rho_{-k+p-1} \\ \rho_{-k+1} & \rho_{-k+2} & \cdots & \rho_{-k+p} \\ \vdots & \vdots & & \\ \rho_{-k+p-1} & \rho_{-k+p} & \cdots & \rho_{-k+2p-2} \end{vmatrix}} \tag{A2}$$

where the denominator is not zero.

In light of (A1) and the fact that the autocorrelation function is an even function of the lag, the numerator of (A2) may be written

$$-\frac{1}{\phi_p}(-1)^p\left(1-\sum_{i=1}^{p}\phi_i\right) H_p\,[\rho_{-k}]\,.$$

for $-k \leq -p -q$, by applying appropriate row and column operations. Then

$$S_p(\rho_{-k}) = -\frac{1}{\phi_p}\, C$$

and $C \neq 0$ since the process is stationary.

Suppose n is the smallest positive integer such that

$$S_n(\rho_{-k}) = C \neq 0$$

and

$$S_n(\rho_{M+1}) \neq C \tag{A3}$$

for $-k \leq M$. Then from (A2)

$$\begin{vmatrix} 1 & 1 & \cdots & 1 \\ \rho_{-k} & \rho_{-k+1} & \cdots & \rho_{-k+n} \\ \vdots & \vdots & & \\ \rho_{-k+n-1} & \rho_{-k+n} & \cdots & \rho_{-k+2n-1} \end{vmatrix} = (-1)^n C \begin{vmatrix} 0 & 0 & \cdots & 0 & 1 \\ \rho_{-k} & \rho_{-k+1} & \cdots & \rho_{-k+n-1} & \rho_{-k+n} \\ \vdots & \vdots & & \vdots & \vdots \\ \rho_{-k+n-1} & \rho_{-k+n} & \cdots & \rho_{-k+2n-2} & \rho_{-k+2n-1} \end{vmatrix}$$

$$\text{so that} \quad \begin{vmatrix} 1 & 1 & & 1 & 1-(-1)^n C \\ \rho_{-k} & \rho_{-k+1} & \cdots & \rho_{-k+n-1} & \rho_{-k+n} \\ \vdots & \vdots & & \vdots & \vdots \\ \rho_{-k+n-1} & \rho_{-k+1} & \cdots & \rho_{-k+2n-2} & \rho_{-k+2n-1} \end{vmatrix} = 0. \quad \text{(A4)}$$

That is a determinant such as (A4) may be exhibited for each k such that $-k \leq M$. Then using the symmetry of ρ, for each k constants $b_1(k)$, $b_2(k)$, ..., $b_p(k)$ exist such that

$$b_1(k)\begin{bmatrix} 1 \\ \rho_k \\ \vdots \\ \rho_{k-n+1} \end{bmatrix} + \ldots + b_n(k)\begin{bmatrix} 1 \\ \rho_{k-n+1} \\ \vdots \\ \rho_{k-2n+2} \end{bmatrix} + \begin{bmatrix} 1-(-1)^n C \\ \rho_{k-n} \\ \vdots \\ \rho_{k-2n+1} \end{bmatrix} = 0.$$

Consequently

$$\begin{bmatrix} 1 & 1 & \cdots & 1 \\ \rho_k & \rho_{k-1} & \cdots & \rho_{k-n+1} \\ \vdots & \vdots & & \vdots \\ \rho_{k-n} & \rho_{k-n-1} & \cdots & \rho_{k-2n+1} \end{bmatrix} \begin{bmatrix} b_1(k)-b_1(k+1) \\ \cdot \\ \cdot \\ \cdot \\ b_n(k)-b_n(k+1) \end{bmatrix} = 0. \quad \text{(A5)}$$

Then either the determinant of the nxn matrix of (A5) is zero or the vector is null. However the determinant of the nxn matrix is the numerator of $S_{n-1}(\rho_m)$. That this quantity would be zero is a contradiction to the assumption that n is the smallest value such that (A3) holds.

Hence

$$b_i\,(k) = b_i\,(k+1) \quad \text{for} \quad i = 1, 2, \ldots, n$$
$$\text{and } k \geq -M\ .$$

Thus the constants $b_i\,(k)$ do not depend on k and we have

$$\rho_{k-n+1} = \sum_{i=1}^{n} \psi_i \rho_{k-n+1-i}$$

for $k-n+1 \geq -M-n+1$. It follows therefore that $n = p$ and $M = -q-p$, completing the proof.

APPENDIX 2

(x,y): x = minimum lag of acf, y = maximum lag of acf required to compute $s_n(k)$, n=1,2,..., k=...,-1,0,1,2, ...

$k \backslash n$	1	2	3	4	5
.	.	.	.	.	.
.	.	.	.	.	.
.	.	.	.	.	.
-12	(11,12)	(9,12)	(7,12)	(5,12)	(3,12)
-11	(10,11)	(8,11)	(6,11)	(4,11)	(2,11)
-10	(9,10)	(7,10)	(5,10)	(3,10)	(1,10)
-9	(8,9)	(6,9)	(4,9)	(2,9)	(0,9)
-8	(7,8)	(5,8)	(3,8)	(1,8)	(0,8)
-7	(6,7)	(4,7)	(2,7)	(0,7)	(0,7)
-6	(5,6)	(3,6)	(1,6)	(0,6)	(0,6)
-5	(4,5)	(2,5)	(0,5)	(0,5)	(0,5)*
-4	(3,4)	(1,4)	(0,4)	(0,4)*	(0,5)*
-3	(2,3)	(0,3)	(0,3)*	(0,4)*	(0,6)
-2	(1,2)	(0,2)*	(0,3)*	(0,5)	(0,7)
-1	(0,1)*	(0,2)*	(0,4)	(0,6)	(0,8)
0	(0,1)*	(0,3)	(0,5)	(0,7)	(0,9)
1	(1,2)	(1,4)	(1,6)	(1,8)	(1,10)
2	(2,3)	(2,5)	(2,7)	(2,9)	(2,11)
3	(3,4)	(3,6)	(3,8)	(3,10)	(3,12)
4	(4,5)	(4,7)	(4,9)	(4,11)	(4,13)
5	(5,6)	(5,8)	(5,10)	(5,12)	(5,14)
.	.	.	.	.	.
.	.	.	.	.	.
.	.	.	.	.	.

(x,y): x = minimum lag of acf, y = maximum lag of acf required to compute $r_n(k)$, n = 1,2,..., k = ...,-1,0,1,...

k \ n	1	2	3	4	5
⋮	⋮	⋮	⋮	⋮	⋮
-12	(12,12)	(10,12)	(8,12)	(6,12)	(4,12)
-11	(11,11)	(9,11)	(7,11)	(5,11)	(3,11)
-10	(10,10)	(8,10)	(6,10)	(4,10)	(2,10)
-9	(9,9)	(7,9)	(5,9)	(3,9)	(1,9)
-8	(8,8)	(6,8)	(4,8)	(2,8)	(0,8)
-7	(7,7)	(5,7)	(3,7)	(1,7)	(0,7)
-6	(6,6)	(4,6)	(2,6)	(0,6)	(0,6)
-5	(5,5)	(3,5)	(1,5)	(0,5)	(0,5)
-4	(4,4)	(2,4)	(0,4)	(0,4)	(0,4)*
-3	(3,3)	(1,3)	(0,3)	(0,3)*	(0,5)
-2	(2,2)	(0,2)	(0,2)*	(0,4)	(0,6)
-1	(1,1)	(0,1)*	(0,3)	(0,5)	(0,7)
0	(0,0)*	(0,2)	(0,4)	(0,6)	(0,8)
1	(1,1)	(1,3)	(1,5)	(1,7)	(1,9)
2	(2,2)	(2,4)	(2,6)	(2,8)	(2,10)
3	(3,3)	(3,5)	(3,7)	(3,9)	(3,11)
4	(4,4)	(4,6)	(4,8)	(4,10)	(4,12)
5	(5,5)	(5,7)	(5,9)	(5,11)	(5,13)
⋮	⋮	⋮	⋮	⋮	⋮

APPENDIX 3

Theorem 2A: Let r_1, r_2 and r_3 be the roots of the characteristic equation $1 - \phi_1 x - \phi_2 x^2 - \phi_3 x^3 = 0$ of an ARMA (3,1) process in which M, $1 \le M \le 3$, of the roots r_1, r_2 and r_3 lie on the unit circle. Let the integer k, $1 \le k \le M$ be the number of distinct roots having modulus 1. Then ρ_m^* as defined in Definition 6 satisfies a linear homogeneous difference equation with constant coefficients whose order is less than or equal to k for m = 0, ±1, ..., and whose characteristic equation has all of its roots distinct and on the unit circle.

Preliminaries to Proof: For a stationary ARMA (3,1) process,

$$\rho_m = \phi_1\rho_{m-1} + \phi_2\rho_{m-2} + \phi_3\rho_{m-3} \text{ for } m \ge 2. \qquad (1)$$

Now the general solution to (1) is

$$\rho_m = A_1 r_1^{-|m|} + A_2 r_2^{-|m|} + A_3 r_3^{-|m|} \quad \text{for 3 distinct roots,} \qquad (2)$$

$$\rho_m = (B_1 + B_2 m) r_1^{-|m|} + B_3 r_2^{-|m|} \quad \text{for } r_1 \text{ repeated once} \qquad (3)$$

and

$$\rho_m = (C_1 + C_2 m + C_3 m^2) r_1^{-|m|} \quad \text{for } r_1 \text{ repeated twice.} \qquad (4)$$

In these equations A_1, A_2, A_3, B_1, B_2, B_3, C_1, C_2, and C_3 are all constants determined by

$$\rho_0 = 1$$

$$\rho_1 = \frac{\phi_1 + \phi_2\phi_3}{1-\phi_2-\phi_3(\phi_1+\phi_3)} + \theta_1 k_1 ,$$

where letting $A = \phi_1 + \phi_2\phi_3 + (\phi_2+\phi_1\phi_3)(\phi_1+\phi_3)$ and

$$B = 1 - \phi_2 - \phi_3(\phi_1+\phi_3) ,$$

$$k_1 = \frac{(\phi_1+\phi_2\phi_3)A - [1-\phi_3^2-\phi_2(\phi_2+\phi_1\phi_3)]B}{B\{B[1-\theta_1(\phi_1-\theta_1)]-\theta_1 A\}} ,$$

and

$$\rho_2 = (\phi_1+\phi_3)\rho_1 + \phi_2 .$$

Now let $G_1 = r_1^{-1}$, $G_2 = r_2^{-1}$ and $G_3 = r_3^{-1}$. Then the constants given in equations (2), (3) and (4) are:

$$A_1 = (G_2-G_3)(\rho_1(G_2+G_3) - \rho_2 - \rho_0 G_2 G_3)/D$$

$$A_2 = (G_3-G_1)(\rho_1(G_3+G_1) - \rho_2 - \rho_0 G_1 G_3)/D.$$

$$A_3 = (G_1-G_2)(\rho_1(G_1+G_2) - \rho_2 - \rho_0 G_1 G_2)/D$$

where $$D = (G_1-G_2)(G_1-G_3)(G_3-G_2) \ ;$$

$$B_1 = \frac{G_3^2 + 2\rho_1 G_1 - 2G_1 G_3 - \rho_2}{(G_1-G_3)^2}$$

$$B_2 = \frac{\rho_2 - \rho_1(G_1+G_3) + G_1 G_3}{G_1(G_1-G_3)^2}$$

$$B_3 = \frac{\rho_2 - 2G_1\rho_1 + G_1^2}{(G_1-G_3)^2} \ ;$$

and

$$C_1 = 1$$

$$C_2 = \frac{4\rho_1 G_1 - \rho_2 - 3G_1^2}{2G_1^2}$$

$$C_3 = \frac{\rho_2 - 2\rho_1 G_1 + G_1^2}{2G_1^2} .$$

Finally recall

$$\phi_1 = G_1 + G_2 + G_3$$

$$\phi_2 = -(G_1 G_2 + G_1 G_3 + G_2 G_3$$

$$\phi_3 = G_1 G_2 G_3 .$$

Essentially, the proof involves expressing the appropriate general solution (2), (3), or (4) of the difference equation (1) in terms of G_1, G_2 and G_3 and taking the limit of the general solution as the specified roots approach the unit circle. In each case, the roots which approach the unit circle are arbitrarily chosen. We have the following results:

Proof of Theorem 2A:

Case 1: $M = 1$, $k = 1$, all roots real. Let $G_1 \to 1$ while G_2 and G_3 remain fixed with $|G_2| < 1$ and $|G_3| < 1$. Then $\lim_{G_1 \to 1} A_1 = 1$ $\lim_{G_1 \to 1} A_2 = 0$ and $\lim_{G_1 \to 1} A_3 = 0$. Also, $\lim_{G_1 \to 1} \rho_1 = 1$ and $\lim_{G_1 \to 1} \rho_2 = 1$.

Thus, the acf of the process is given by $\rho_m^* = 1$ for $m = 0, \pm 1, \ldots$; i.e., $\rho_m^* = \rho_{m-1}^*$ and we see that ρ_m^* is a solution to a first order difference equation.

Case 2: $M = 2$, $k = 1$, all roots real. Assume G_1 has multiplicity 2 and take the limits as $G_1 \to 1$ where $|G_3| < 1$ with G_3 fixed. The results are identical to the results given in Case 1; i.e., $\rho_m^* = \rho_{m-1}^*$ for $m = 0, \pm 1, \ldots$.

Case 3: $M = 3$, $k = 1$, all roots real. Assuming G_1 has multiplicity 3, these results are also identical to those in Case 1.

Case 4: $M = 2$, $k = 2$, 2 complex roots, 1 real root. Let $G_1 = x + jy = de^{j\theta}$ and $G_2 = G_1 = x - jy = de^{-j\theta}$ where $d = \sqrt{x^2 + y^2}$, $j = \sqrt{-1}$ and $\tan \theta = x/y$. Assume G_3 is fixed with $|G_3| < 1$. The limits are taken as $d \to 1^-$. Then, $\lim_{d \to 1^-} A_1 = \lim_{d \to 1^-} A_2 = 1/2$, $\lim_{d \to 1^-} A_3 = 0$, $\lim_{d \to 1^-} \rho_1 = \cos \theta$, and $\lim_{d \to 1^-} \rho_2 = \cos 2\theta$. Thus, the general solution for the difference equation becomes $\rho_m^* = \frac{1}{2} e^{j\theta|m|} + \frac{1}{2} e^{-j\theta|m|} = \cos \theta m$ for $m = 0, \pm 1, \ldots$, which is the solution to the second order difference equation whose characteristic equation has roots $e^{j\theta}$ and $e^{-j\theta}$.

Case 5: $M = 2$, $k = 2$, all roots real. Here, we have $G_1 \to 1$, $G_2 \to -1$ with G_3 fixed and $|G_3| < 1$. The limits are taken with the restriction $G_2 = -G_1$. Then $\lim_{G_1 \to 1} A_1 = \frac{1}{2}(\rho_1^* + 1)$, $\lim_{G_1 \to 1} A_2 = \frac{1}{2}(1 - \rho_1^*)$, $\lim_{G_1 \to 1} A_3 = 0$, where

$$\rho_1^* = \lim_{G_1 \to 1} \rho_1 = \frac{2G_3}{1+G_3^2} - \frac{2\theta_1(1-G_3^2)^2}{(1+G_3^2)[(1-\theta_1(G_3-\theta_1))(1+G_3^2)-G_3\theta_1(3-G_3^2)]} .$$

Also, $\lim_{G_1 \to 1} \rho_2 = 1$. Thus, the general solution to the difference equation becomes

$$\rho_m^* = \frac{1}{2}[(\rho_1^*+1) + (-1)^{|m|}(1-\rho_1^*)]$$

$$= \begin{cases} \rho_1^* & , m \text{ odd} \\ 1 & , m \text{ even} \end{cases} ;$$

i.e., $\rho_m^* = \rho_{m-2}^*$ for $M = 0, \pm 1, \ldots$, which implies that ρ_m^* satisfies a second order difference equation with roots ± 1.

<u>Case 6</u>: $M = 3$, $m = 2$, all roots real. Let G_1 have multiplicity 2, $G_1 \to 1$, and $G_3 \to -1$. The limits are taken with the restriction $G_3 = -G_1$. Then $\lim_{G_1 \to 1} B_1 = 1$, $\lim_{G_1 \to 1} B_2 = \lim_{G_1 \to 1} B_3 = 0$,

$\lim_{G_1 \to 1} \rho_1 = \lim_{G_1 \to 1} \rho_2 = 1$. These results are seen to be identical to those in Case 1; i.e., $\rho_m^* = \rho_{m-1}^*$ for $m = 0, \pm 1, \ldots$. (This case is distinct in that the order of the resulting difference equation is less than m, the number of distinct roots on the circle, while in the preceding cases, the order was equal to m.)

<u>Case 7</u>: $M = 3$, $m = 3$, 2 complex roots, 1 real root. Let $G_1 = de^{j\theta}$, $G_2 = de^{-j\theta}$, and take the limits as $|G_1| \to 1$, $|G_2| \to 1$ and $G_3 \to 1$. Under the restriction $|G_1| = |G_2| = |G_3|$, we let $d = G_3$ and take the limits as $G_3 \to 1$. Then:

$$\lim_{G_3 \to 1} A_1 = \lim_{G_3 \to 1} A_2 \frac{1}{2}\left(\frac{1}{2+\cos\theta}\right), \quad \lim_{G_3 \to 1} A_3 = \frac{1+\cos\theta}{2+\cos\theta} ,$$

$$\lim_{G_3 \to 1} \rho_1 = \frac{2\cos\theta+1}{\cos\theta+2} , \quad \lim_{G_3 \to 1} \rho_2 = \frac{\cos\theta(2\cos\theta+1)}{2+\cos\theta} .$$ Thus, ρ_m^* is a solution to a third order difference equation for $m = 0 \mp 1, \mp 2 \ldots$.

Other cases are possible but have been omitted here because of their near equivalence to one of the previous cases. For example if $M = 1$, $k = 1$ and $G_1 \to -1$ the result is the same as Case 1 except $\rho_m^* = -\rho_{m-1}^*$.

```
      SUBROUTINE RANDS(MXROW,MROW,NEG,KOL,IRHO,G,R,S,IWK,IRS)
C
C     ..................................................................
C
C     DESCRIPTION OF PARAMETERS
C       MXROW - ROW DIMENSION OF R AND S ARRAYS IN CALLING PROGRAM.
C               (INPUT)
C        MROW - NUMBER OF ROWS OF R AND S ARRAYS TO BE CALCULATED (INPUT)
C         NEG - NUMBER OF ROWS OF NEGATIVE LAG IN R AND S ARRAYS
C                    NEG  <  MROW - 1
C               (INPUT)
C         KOL - NUMBER OF COLUMNS OF R AND S ARRAYS TO BE CALCULATED
C               R AND S ARRAYS MUST BE DIMENSIONED TO AT LEAST KOL COLUMNS
C               IN CALLING PROGRAM.  (INPUT)
C                 KOL  <  (MROW + 2)/2  AND  0  <  KOL  <  12
C               THE RESTRICTION THAT KOL IS LESS THAN TWELVE IS IMPOSED BY
C               THE WRITE STATEMENTS.  WITH IRS .NE. 0, KOL MAY BE AS
C               LARGE AS PERMITTED BY MROW.
C        IRHO - IRHO .EQ. 0, R AND S CALCULATED FOR AUTOCORRELATIONS IN
C               COLUMN 1 OF THE R ARRAY.
C               IRHO .NE. 0, R AND S CALCULATED FOR (-1)**LAG TIMES THE
C               AUTOCORRELATION FUNCTION IN COLUMN 1 OF THE R ARRAY.
C               (INPUT)
C           G - ARRAY OF AUTOCORRELATION ESTIMATES OF LAG ZERO TO N FOR
C               THE FIRST COLUMN OF THE R ARRAY.  WHERE N IS AT LEAST THE
C               MAXIMUM OF   MROW-NEG-1  AND  NEG .  (INPUT)
C           R - R ARRAY.  DIMENSIONED MXROW BY K IN CALLING PROGRAM,
C               WHERE K IS GREATER THAN OR EQUAL TO KOL.  (OUTPUT)
C           S - S ARRAY.  DIMENSIONED AS R ARRAY IN CALLING PROGRAM.
C               (OUTPUT)
C         IWK - WORK VECTOR DIMENSIONED AT LEAST KOL IN CALLING PROGRAM.
C         IRS - PRINT CONTROL PARAMETER.  (INPUT)
C               IRS .EQ. 0, R AND S ARRAYS ARE PRINTED ON UNIT NUMBER IPR,
C               WHICH IS SET IN THE DATA STATEMENT.
C               IRS .NE. 0, CAUSES PRINTING TO BE SUPPRESSED.
C
C     ..................................................................
C
      DIMENSION R(MXROW,1),S(MXROW,1),G(1),IWK(1)
      DATA IBLK/1H /,ISTR/1H*/,IPR/6/,IZRO/0/
      KOLM1=KOL-1
      DO 5 IC=1,KOL
      DO 5 IR=1,MROW
      R(IR,IC)=0.0
    5 S(IR,IC)=0.0
      MN=MROW-NEG
      DO 10 I=1,MN
      NEGI=NEG+I
      S(NEGI,1)=1.0
   10 R(NEGI,1)=G(I)
      DO 20 I=1,NEG
      NEGI=NEG+2-I
      S(I,1)=1.0
   20 R(I,1)=G(NEGI)
```

```
      IF(IRHO .EQ. 0) GO TO 40
      DO 30 I=1,MROW
      NEGI=I-NEG-1
   30 R(I,1)=R(I,1)*(-1)**(NEGI)
   40 J1=MROW
      DO 70 IC=2,KOL
      ICM1=IC-1
      J1=J1-1
      DO 60 IR=1,J1
      IRP1=IR+1
      IF(ABS(R(IR,ICM1)) .EQ. 0.0) GO TO 50
      S(IR,IC)=S(IRP1,ICM1)*(R(IRP1,ICM1)/R(IR,ICM1)-1.0)
      GO TO 60
   50 S(IR,IC)=-S(IRP1,ICM1)
   60 CONTINUE
      J1=J1-1
      DO 70 IR=1,J1
      IRP1=IR+1
   70 R(IR,IC)=R(IRP1,ICM1)*(S(IRP1,IC)/S(IR,IC)-1.0)
      IF(IRS .NE. 0) RETURN
      IF(IRHO .EQ. 0) GO TO 80
      WRITE(IPR,200)
      GO TO 90
   80 WRITE(IPR,210)
   90 WRITE(IPR,220) (IC,IC=1,KOL)
      DO 110 IR=1,MROW
      IRN=IR-NEG-1
      DO 100 IC=1,KOL
      IWK(IC)=IBLK
  100 IF(IR .EQ. NEG+2-IC) IWK(IC)=ISTR
  110 WRITE(IPR,230) IRN,     (R(IR,K),IWK(K),K=1,KOL)
      WRITE(IPR,240) IZRO,(IC,IC=1,KOLM1)
      DO 130 IR=1,MROW
      IRN=IR-NEG-1
      DO 120 IC=1,KOL
      IWK(IC)=IBLK
      IF(IR .EQ. NEG+2-IC) IWK(IC)=ISTR
  120 IF(IR .EQ. NEG+3-IC) IWK(IC)=ISTR
  130 WRITE(IPR,230) IRN,     (S(IR,K),IWK(K),K=1,KOL)
      RETURN
  200 FORMAT(//20X,27H( F(M) = ((-1)**M)*ACF(M) ))
  210 FORMAT(//25X,17H( F(M) = ACF(M) ))
  220 FORMAT(//30X,7HR-ARRAY//3H  K,11(9X,I2))
  230 FORMAT(2H (,I3,1H),11(1X,F9.4,A1))
  240 FORMAT(//30X,7HS-ARRAY//3H  K,11(9X,I2))
      END
```

```
      SUBROUTINE DSTAT(MXROW,MROW,NEG,R,S,IP,IQ,IPNT)
C
C     ..............................................................
C
C     DESCRIPTION OF PARAMETERS
C        MXROW - ROW DIMENSION OF R AND S ARRAYS IN CALLING PROGRAM.
C                (INPUT)
C         MROW - NUMBER OF ROWS TO WHICH R AND S WERE CALCULATED.  (INPUT)
C          NEG - NUMBER OF ROWS OF NEGATIVE LAG IN R AND S ARRAYS
C                (INPUT)
C            R - R ARRAY.  DIMENSIONED MXROW BY K IN CALLING PROGRAM,
C                WHERE K IS GREATER THAN OR EQUAL TO KOL.  (INPUT)
C            S - S ARRAY.  DIMENSIONED AS R ARRAY IN CALLING PROGRAM.
C                (INPUT)
C           IP - (INPUT) MAXIMUM ORDER OF AUTOREGRESSION PERMITTED.
C                (OUTPUT) ORDER OF AUTOREGRESSION SELECTED.
C           IQ - (INPUT) MAXIMUM ORDER OF MOVING AVERAGE PERMITTED.
C                (OUTPUT) ORDER OF MOVING AVERAGE SELECTED.
C         IPNT - PRINT CONTROL PAREMETER  (INPUT)
C                IPNT .EQ. 0, D STATISTIC VALUES ARE PRINTED.
C                IPNT .NE. 0, PRINTING IS SUPPPRESSED.
C                WITH IPNT = 0 THE WRITE STATEMENTS REQUIRE THE INPUT VALUE
C                OF IQ TO BE LESS THAN TEN.
C                PRINTING IS ON UNIT IPR WHICH IS SET IN THE DATA STATEMENT.
C
C     NOTE
C             NEG MUST BE GREATER THAN OR EQUAL TO  IP+IQ+3, FOR IP AND IQ
C        THEIR INPUT VALUES.  FURTHERMORE MROW MUST BE GREATER THAN OR EQUAL
C        TO  NEG+IP+IQ+4, FOR IP AND IQ EQUAL TO THEIR INPUT VALUES.
C             IF (IP+1)*(IQ+1) IS GREATER THAN 50 (WHERE IP AND IQ ARE AT
C        THEIR INPUT VALUES) THEN THE DIMENSION OF STAT MUST BE CHANGED
C        ACCORDINGLY.
C
C     ..............................................................
C
      DIMENSION R(MXROW,1),S(MXROW,1),STAT(50)
      DATA IPR/6/,IZRO/0/
      JPP=IP+1
      JQQ=IQ+1
      IOP=JPP*JQQ
      DO 40 I=1,JPP
      JP=I-1
      DO 40 J=1,JQQ
      JQ=J-1
      JK=JQQ*(I-1)+J
      ISUB2=NEG-JP+JQ+2
      ISUB1=ISUB2-1
      IP1=I+1
      IM1=I-1
      BOT=(S(ISUB2,I)+S(ISUB1,IP1))**2
      IF(I .GT. 1) GO TO 20
      DO 10 JI=1,3
      NEGM=NEG-JP-JQ-JI+1
      NEGP=NEG-JP+JQ+JI+1
```

```
10 BOT=BOT+R(NEGM,I)**2 + R(NEGP,I)**2
   GO TO 35
20 ISUB3=ISUB2+1
   BOT=BOT/(S(ISUB3,IM1) + S(ISUB2,I))**2
   DO 30 JI=1,3
   NEGM=NEG-JP-JQ-JI+1
   NEGP=NEG-JP+JQ+JI+1
   NEGM1=NEGM+1
   NEGP1=NEGP+1
30 BOT=BOT+(R(NEGM,I)/R(NEGM1,IM1))**2+(R(NEGP,I)/R(NEGP1,IM1))**2
35 NEG0=NEG-JP-JQ
   NEG1=NEG0+1
   ARG=S(NEG0,IP1)/(BOT*S(NEG1,I))
40 STAT(JK)=ABS(ARG)
   IMX=1
   XMX=STAT(1)
   DO 50 I=1,IOP
   IF(STAT(I) .LT. XMX) GO TO 50
   IMX=I
   XMX=STAT(I)
50 CONTINUE
   IP=(IMX-1)/JQQ
   IQ=IMX-(IP*JQQ)-1
   IF(IPNT .NE. 0) RETURN
   WRITE(IPR,100)
   JQQ=JQQ-1
   WRITE(IPR,110) IZRO,(I,I=1,JQQ)
   DO 60 J=1,JPP
   JM1=J-1
   IST=JM1*(JQQ+1)+1
   IEND=IST+JQQ
60 WRITE(IPR,120) JM1,(STAT(I),I=IST,IEND)
   WRITE(IPR,130) IP,IQ
   RETURN
100 FORMAT(///50X,11HD STATISTIC//29X,11HORDER OF MA)
110 FORMAT(13H ORDER OF AR ,10(4X,I3,4X))
120 FORMAT(/4X,I2,6X,10(1X,E10.4))
130 FORMAT(/34H  ORDER OF AUTOREGRESSION SELECTED,2X,I4/
  1         34H  ORDER OF MOVING AVERAGE SELECTED,2X,I4)
   END
```

REFERENCES

1. Akaike, H. "Statistical Predictor Identification". Annals of the Institute of Statistical Mathematics 22 (1970): 203-217.

2. Akaike, H. "A New Look at the Statistical Model Identification." IEEE Transactions on Automatic Control AC-19 (1974) 716-723.

3. Anderson, O.D. "A Box-Jenkins Analysis of the Coloured Fox Data from Nain, Labrador". The Statistician 26, No.1 (1977): 51-73.

4. Anderson, T.W. The Statistical Analysis of Time Series. New York: John Wiley and Sons, Inc. 1971.

5. Box, G.E.P. and Jenkins, G.M. Time Series Analysis: Forecasting and Control. San Francisco: Holden-Day, 1970.

6. Gray, H.L., Atchison, T.A. and McWilliams, G.V. (1971). "Higher Order G-Transformations". SIAM J. Numer.Anal.8, (2), 365-381.

7. Gray, H.L., Houston, A.G. and Morgan, F.W., "On G-Spectral Estimation". Proceedings of the 1976 Tulsa Symposium on Applied Time Series. New York: Academic Press, to appear 1978.

8. Kelley, G.D. "A New Approach to ARMA Modeling, the Non-stationary Case". Ph.D. Dissertation, Southern Methodist University, 1977.

9. McIntire, D.D. "A New Approach to ARMA Modeling". Ph.D. Dissertation, Southern Methodist University, 1977.

10. Pye, W.C. and Atchison, T.A. "An Algorithm for the Computation of the Higher Order G-Transformation". SIAM J. Numer. Anal. 10, No.1 (1973) 1-7.

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER (14) TR-126	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) (6) A New Approach to ARMA Modeling.		5. TYPE OF REPORT & PERIOD COVERED (9) Technical Rep[redacted]t,
		6. PERFORMING ORG. REPORT NUMBER 126
7. AUTHOR(s) (10) H. L./Gray, G. D./Kelley, D. D./McIntire		8. CONTRACT OR GRANT NUMBER(s) (15) N00014-75-C-0439 N00014-C-75-0439
9. PERFORMING ORGANIZATION NAME AND ADDRESS Southern Methodist University Dallas, Texas 75275		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS Project NR 042-280
11. CONTROLLING OFFICE NAME AND ADDRESS Probability and Statistics Program Office of Naval Research Arlington, VA 22217	(11) 1	12. REPORT DATE Dec[redacted]77
		13. NUMBER OF PAGES ~~80~~ (12) 82p.
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office)		15. SECURITY CLASS. (of this report) UNCLASSIFIED
		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

This document has been approved for public release and sale; its distribution is unlimited. Reproduction in whole or in part is permitted for any purpose of the United States government.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Autoregressive - Moving Average Processes
G-Spectral Estimation
G-Transforms
Forecasting

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

In recent years the Box-Jenkins method has become a popular technique for forecasting future behavior of a time series. Once the forecast model is known the method is very easy to employ and adequate computer packages are available for most purposes. Unfortunately the problem of determining the appropriate forecast model has, for models of any complexity, been one of the major stumbling blocks to the user of this method.

In this paper a satisfactory solution to that problem is obtained and it is demonstrated by numerous examples how this greatly enlarges the class of data

next page

404 198

Jmc

20.

sets which can be adequately modeled by autoregressive-moving average models. This new approach is sufficiently unequivocal that most users will find it easy to implement.